(19)日本国特許庁（ＪＰ）　(12) **公開特許公報**（Ａ）　(11)特許出願公開番号
特開2002－324000
（P2002－324000A）
(43)公開日　平成14年11月８日(2002.11.8)

(51)Int.Cl.[7]	識別記号	ＦＩ	ﾃｰﾏｺｰﾄﾞ(参考)
Ｇ０６Ｆ　12/00	５１４	Ｇ０６Ｆ　12/00	５１４Ｅ　５Ｂ０６５
	５３１		５３１Ｄ　５Ｂ０８２
	５４５		５４５Ａ
3/06	３０４	3/06	３０４Ｆ

審査請求　未請求　請求項の数17　ＯＬ　外国語出願　（全 54 頁）

(21)出願番号　特願2002－3193(P2002－3193)
(22)出願日　平成14年１月10日(2002.1.10)
(31)優先権主張番号　０９／７６０３４４
(32)優先日　平成13年１月12日(2001.1.12)
(33)優先権主張国　米国（ＵＳ）

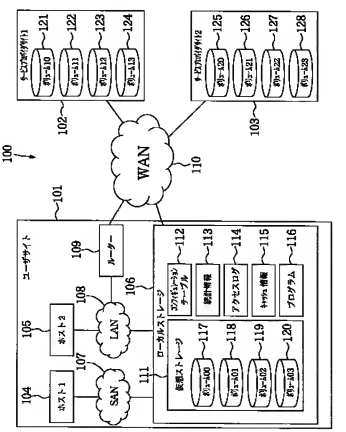
(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　渡辺　直企
アメリカ合衆国　カリフォルニア州　プリスベーン　シエラポイントパークウェイ　2000　日立アメリカ　リサーチアンドディベロップメント部内
(72)発明者　山本　彰
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(74)代理人　100091096
弁理士　平木　祐輔

最終頁に続く

(54)【発明の名称】　仮想ボリューム・ストレージ

(57)【要約】
【課題】　データ・ストレージ・サービスを提供するシステムを提供する。
【解決手段】　データ・ストレージ・サービスを提供するように構成されているサービス・プロバイダ・サイト１０２，１０３及びサービス・プロバイダ・サイトにＷＡＮ１１０により結合されたユーザ・サイト１０１を備える。ユーザ・サイトは仮想ストレージ１１１を持つローカル・ストレージ１０６を備え、仮想ストレージは同期ボリューム及び非同期ボリュームを備え、ローカル・ストレージは同期ボリューム内に書き込まれたデータをサービス・プロバイダ・サイトに即座に送信し、非同期ボリュームに書き込まれたデータを所定のスケジュールでサービス・プロバイダ・サイトに送信し、データがローカル・ストレージに格納されていない場合にサービス・プロバイダ・サイトからデータを読み込むように構成されている。

図　1



【特許請求の範囲】
【請求項１】　データ・ストレージ・サービスを提供するシステムであって、
データ・ストレージ・サービスを提供するように構成されているサービス・プロバイダと、
ワイド・エリア・ネットワーク（ＷＡＮ）によりサービス・プロバイダ・サイトに結合されたユーザ・サイトを備え、前記ユーザ・サイトは仮想ストレージを持つローカル・ストレージを備え、前記仮想ストレージは同期ボリューム及び非同期ボリュームを備え、前記ローカル・ストレージは前記同期ボリューム内に書き込まれたデータを前記サービス・プロバイダ・サイトに即座に送信し、前記非同期ボリュームに書き込まれたデータを所定のスケジュールで前記サービス・プロバイダ・サイトに送信し、データが前記ローカル・ストレージに格納されていない場合に前記サービス・プロバイダ・サイトからデータを読み込むように構成されているシステム。
【請求項２】　前記ローカル・ストレージがホストによる当該ローカル・ストレージへのアクセス活動を記録するように構成されていることを特徴とする請求項１記載のシステム。
【請求項３】　前記ローカル・ストレージが前記仮想ストレージ内のボリュームへのアクセス活動を記録することを特徴とする請求項１記載のシステム。
【請求項４】　前記ローカル・ストレージが前記仮想ストレージ内のボリュームのサブエリアでのアクセス活動を記録することを特徴とする請求項１記載のシステム。
【請求項５】　分析のため前記サービス・プロバイダ・サイトでアクセス活動を行えることを特徴とする請求項２記載のシステム。
【請求項６】　データ・ストレージ・サービスを提供する方法であって、
同期ボリュームと非同期ボリュームとを有する仮想ストレージを備えるローカル・ストレージを提供するステップと、
選択的に実行するステップであって、
前記仮想ストレージの同期ボリューム内に書き込まれたデータをサービス・プロバイダ・サイトに即座に送信するステップと、
前記仮想ストレージの非同期ボリューム内に書き込まれたデータを前記サービス・プロバイダ・サイトに所定のスケジュールで送信するステップと、
データが前記ローカル・ストレージ内に格納されていない場合に前記サービス・プロバイダ・サイトからデータを読み込むステップのうち少なくとも１つを選択的に実行するステップとを含む方法。
【請求項７】　前記ローカル・ストレージへのアクセス活動を記録するステップを更に含むことを特徴とする請求項６記載の方法。
【請求項８】　前記サービス・プロバイダ・サイトへのアクセス活動の記録を提供するステップを更に含むことを特徴とする請求項７記載の方法。
【請求項９】　仮想ストレージの同期ボリューム内に書き込まれたデータをサービス・プロバイダ・サイトに即座に送信するステップ、
前記仮想ストレージの非同期ボリューム内に書き込まれたデータを前記サービス・プロバイダ・サイトに所定のスケジュールで送信するステップ、及びデータが前記仮想ストレージを含むローカル・ストレージ内に格納されていない場合に前記サービス・プロバイダ・サイトからデータを読み込むステップのうち少なくとも１つを選択的に実行するステップを含むデータ・ストレージ・サービスを提供する方法をコンピュータに実行させるためのプログラムを格納する電子的に読み取り可能な媒体。
【請求項１０】　仮想ストレージの同期ボリューム内に書き込まれたデータをサービス・プロバイダ・サイトに即座に送信するステップ、
前記仮想ストレージの非同期ボリューム内に書き込まれたデータを前記サービス・プロバイダ・サイトに所定のスケジュールで送信するステップ、及びデータが前記仮想ストレージを含むローカル・ストレージ内に格納されていない場合に前記サービス・プロバイダ・サイトからデータを読み込むステップのうち少なくとも１つを選択的に実行するステップを含むデータ・ストレージ・サービスを提供する方法をコンピュータに実行させるための搬送波に埋め込まれたプログラム・コード。
【請求項１１】　データ・ストレージ・システム内のデータを処理する装置であって、
仮想ストレージを持つローカル・ストレージを備えるユーザ・サイトであって、前記仮想ストレージは同期領域及び非同期領域を含み、前記ローカル・ストレージは前記同期領域内に書き込まれたデータをサービス・プロバイダ・サイトに即座に送信し、前記非同期領域に書き込まれたデータを所定のスケジュールで前記サービス・プロバイダ・サイトに送信し、データが前記ローカル・ストレージに格納されていない場合に前記サービス・プロバイダ・サイトからデータを読み込むように構成されているユーザ・サイトを含むことを特徴とする装置。
【請求項１２】　ローカル・ストレージへのコマンド信号に応答して、
仮想ストレージ内の同期領域に書き込まれたデータをサービス・プロバイダ・サイトに即座に送信するステップ、
前記仮想ストレージ内の非同期領域に書き込まれたデータを前記サービス・プロバイダ・サイトに所定のスケジュールで送信するステップ、
データが前記ローカル・ストレージ内に格納されていない場合に前記サービス・プロバイダ・サイトからデータを読み込むステップのうち少なくとも１つを選択的に実行するデータ・ストレージ・システム内のデータを処理

する方法。
【請求項１３】　データ・ストレージ・システム内のデータを読み込み方法であって、
ホストからローカル・ストレージへ読み込みコマンドを受信するステップと、
前記読み込みコマンドで要求されたデータのボリューム・タイプを判別するステップと、
前記ボリューム・タイプが静的である場合に、前記ローカル・ストレージ内のローカル・ディスクからデータを読み込むステップと、
前記ボリューム・タイプがキャッシュの場合に、データが前記ローカル・ディスクに格納されているかどうかを調べるステップと、
前記データが前記ローカル・ディスクに格納されている場合に、前記ローカル・ディスクからデータを読み込むステップと、
前記データが前記ローカル・ディスクに格納されていない場合に、リモート・サービス・プロバイダ・サイトからデータを読み込み、前記データをローカル・ディスクに格納し、前記ローカル・ディスクから前記データを読み込むステップと、
前記ローカル・ディスクからデータを読み込んだ後に、前記読み込みコマンドに関係する統計情報を更新するステップと、
前記データを前記ホストに返すステップとを含む方法。
【請求項１４】　データ・ストレージ・システム内のデータを処理する装置であって、
仮想ストレージ内の同期領域に書き込まれたデータを即座に受信し、前記仮想ストレージ内の非同期領域に書き込まれたデータを所定のスケジュールで受信し、データがローカル・ストレージに格納されていない場合に前記仮想ストレージを含む前記ローカル・ストレージにデータを送信するように構成されているサービス・プロバイダ・サイトを備える装置。
【請求項１５】　仮想ストレージ内の同期領域に書き込まれたデータを即座に受信するステップ、
前記仮想ストレージ内の非同期領域に書き込まれたデータを所定のスケジュールで受信するステップ、及びデータがローカル・ストレージ内に格納されていない場合に前記仮想ストレージを含む前記ローカル・ストレージにデータを送信するステップのうち１つを選択的に実行するデータ・ストレージ・システム内のデータを処理する方法。
【請求項１６】　データ・ストレージ・システム内にデータを書き込む方法であって、
ホストからローカル・ストレージへの書き込みコマンドを受信するステップと、
前記書き込みコマンドのデータのボリューム・タイプを判別するステップと、
前記ボリューム・タイプが静的である場合に、前記ローカル・ストレージ内のローカル・ディスクに前記データを書き込むステップと、
前記ボリューム・タイプがキャッシュの場合に、前記データが前記ローカル・ディスクに格納されているかどうかを調べるステップと、
前記データが前記ローカル・ディスクに格納されている場合に、キャッシュから前記ローカル・ディスクに前記データを書き込むステップと、
前記データが前記ローカル・ディスクに格納されていない場合に、前記ローカル・ディスクと前記キャッシュ内にデータ領域を割り当てて、前記データを前記キャッシュから前記ローカル・ディスクに書き込むステップと、
前記データのボリューム・タイプが同期かどうかを調べるステップと、
前記ボリューム・タイプが同期の場合に、前記データをリモート・サービス・プロバイダ・サイトと即座に同期させ、前記書き込みコマンドに関係する統計情報を更新するステップと、
前記ボリューム・タイプが同期でない場合に、前記データをリモート・サービス・プロバイダ・サイトと所定のスケジュールに基づいて同期させ、前記書き込みコマンドに関係する統計情報を更新するステップとを含む方法。
【請求項１７】　データ・ストレージ・サービスを提供するシステムであって、
仮想ストレージの同期ボリューム内に書き込まれたデータをサービス・プロバイダ・サイトに即座に送信、
前記仮想ストレージの非同期ボリューム内に書き込まれたデータを前記サービス・プロバイダ・サイトに所定のスケジュールで送信、及びデータが前記仮想ストレージを含むローカル・ストレージ内に格納されていない場合にサービス・プロバイダ・サイトからのデータ読み込みのうち少なくとも１つを選択的に実行する手段を備えるシステム。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明はデータ・ストレージ・システムに関し、より詳細には、ユーザ・サイトに仮想ボリューム・データ・ストレージを備えるシステムと方法に関する。
【０００２】
【従来の技術】大規模なデータ・ストレージ・システムを管理することは非常に困難である。通常、データ・ストレージ・サービス・プロバイダは、ストレージ・ボリューム、データ・バックアップなどを実施するストレージ関連サービスを提供する。ユーザ・サイトから、ユーザがストレージ・サービス・プロバイダのディレクトリ経由でこのストレージに接続する場合、ユーザは、長距離接続を利用してこのストレージに接続する必要がある。このプロセスは、ユーザがローカル・ストレージに

接続する場合に比べて遅延時間が長くなる。

【０００３】World Wide Web（WWW）は、広い地域でストレージ・システムとして効果的な働きをする。ユーザは、ユーザ・サイトにプロキシ・サーバを配備することができる。当業者には周知のように、プロキシ・サーバはクライアント・アプリケーション（Ｗｅｂブラウザなど）とリモート・サーバとの間に置かれるサーバである。プロキシ・サーバは、リモート・サーバ上で利用できるアイテムをキャッシュする機能を備える。プロキシ・サーバは、リモート・サーバに送られるすべての要求を横取りし、プロキシ・サーバ側でその要求を遂行できるかどうかを調べる。プロキシ・サーバ側でその要求を遂行できなければ、プロキシ・サーバはその要求をリモート・サーバに転送し、リモート・サーバに処理させる。プロキシ・サーバは、リモート・サーバ又はサイトから送られたキャッシュ・データを読み込むことができるだけであって、リモート・サーバ又はサイトへの書き込み手順をサポートしているわけではない。さらに、プロキシ・サーバでは、データの特徴に依存するサービスを提供することもできない。そのため、プロキシ・サーバでは、資源を効果的に利用することができず、またプロキシ・サーバを使用した場合、ローカル・ストレージを使用した場合に比べて遅延時間が長くなる。

【０００４】米国特許第5,155,835号では、マルチレベルの階層型動的マッピング・データ・ストレージ・サブシステムを開示している。この特許では、ワイド・エリア・ネットワーク（WAN）間のストレージ・システムを開示しておらず、アクセス・インタフェース・タイプ（ブロック又はファイルといったもの）、目的（ユーザ・データ又はシステム・ファイル又はバックアップ・ファイルといったもの）などのデータの特徴を無視している。したがって、この引用で開示されているサブシステムは不効率である。

【０００５】インターネット・プロトコル小型コンピュータ用周辺機器インタフェース（ｉＳＣＳＩ）では、ネットワーキング・インフラにインターネット・プロトコル（ＩＰ）を利用し、既存のローカル・エリア・ネットワーク及び／又はワイド・エリア・ネットワーク上で大量のブロック・ストレージ（ＳＣＳＩ）データを素早く転送する。ｉＳＣＳＩ（ＩＰ　SAN）では、すべての主要ネットワーキング・プロトコルをサポートできるため、企業全体のネットワーク・アーキテクチャを統一でき、それにより、全体のネットワーク・コスト及び複雑度を低減できる。信頼性を保証するために、ｉＳＣＳＩではＩＰネットワーク用に開発された周知のネットワーク管理ツール及びユーティリティを使用できる。たとえば、ｉＳＣＳＩプロトコルは、Ｗｅｂサイト（http://www.ece.cmu.edu/ips/index.html）のＩＰストレージ・セクション（http://www.ece.cmu.edu/ips/Docs/docs.html）で説明されている。

【０００６】ＩＥＴＦ（Internet Engineering Task Force）の作業グループはhttp://www.ietf.orgで、ネットワーク・ファイル・システム（ＮＦＳ）バージョン３（ＩＥＴＦのｒｆｃ１８１３）及び共通のインターネット・ファイル・システム（ＣＩＦＳ）プロトコル（http://www.cifs.org）を提供している。

【０００７】ＮＦＳは、オープン・オペレーティング・システムであり、このＯＳを利用すればすべてのネットワーク・ユーザが異なる種類のコンピュータに格納されている共有ファイルにアクセスできる。ＮＦＳでは、伝送制御プロトコル／インターネット・プロトコル（ＴＣＰ／ＩＰ）の上位で実行される仮想ファイル・システム（ＶＦＳ）と呼ばれるインタフェースを通じて共有ファイルにアクセスできる。ＮＦＳを使用すると、ネットワークに接続されているコンピュータをクライアントとして動作させながらリモート・ファイルにアクセスし、サーバとして動作させながらリモート・ユーザによるローカル共有ファイルへのアクセスを可能にできる。

【０００８】ＣＩＦＳプロトコルは、数百万台のコンピュータに一度にリモート・ファイルにアクセスできるようにする標準を定めるものである。ＣＩＦＳでは、プラットフォームもコンピュータも異なるユーザでも、新規ソフトウェアをインストールすることなくファイルを共有できる。ＣＩＦＳは、ＴＣＰ／ＩＰ上で実行されるが、ファイル及びプリンタ・アクセス用にMicrosoft Windows（登録商標）にあるサーバ・メッセージ・ブロック（ＳＭＢ）プロトコルを使用する。したがって、ＣＩＦＳでは、すべてのアプリケーション（Ｗｅｂブラウザを含む）がインターネット上のファイルを開き、共有することができる。

【０００９】Ｗｅｂサイトhttp://www.cnt.com及びhttp://www.san.comには、ストレージ又はストレージ・エリア・ネットワーク（ＳＡＮ）にワイド・エリア・ネットワークを接続する方法についての説明がある。WAN接続は、非同期転送モード（ＡＴＭ）、同期光ネットワーク（ＳＯＮＥＴ）などが考えられる。上記の引用では、ストレージ・システムとホスト・デバイスとの間の階層型管理手法を実現する方法を開示していない。

【００１０】従来の方法とシステムの上述の欠陥を克服するシステムと方法が必要である。さらに、アクセス・サービス・プロバイダが信頼できるストレージ・システムを備えることができ、またユーザがストレージ・システムに素早くアクセスできるシステムと方法も必要である。さらに、アクセス・サービス・プロバイダがユーザ統計データ及びユーザ・ログ・データに基づいてローカル・ストレージ・システムをチューニングできるシステムと方法も必要である。

【００１１】

【発明が解決しようとする課題】本発明は、ユーザ・サイトに信頼性の高い仮想ローカル・ストレージを使いや

すい形で実現することができる。本発明はさらに、ユーザ向けの高速なアクセスを可能にする仮想ローカル・ストレージを使いやすい形で実現することができる。本発明はさらに、ローカル・ストレージで追跡されるユーザ統計データ及びユーザ・ログ・データを使用してローカル・ストレージをサービス・プロバイダ側で使いやすい形でチューンアップできるようにし、サービス・プロバイダはこのようなチューニング・サービスについてユーザに課金できる。

【００１２】

【課題を解決するための手段】本発明の一実施形態では、サービス・プロバイダはワイド・エリア・ネットワーク経由でデータ・ストレージ・サービスを提供することができる。仮想ボリューム・ストレージはユーザ・サイトに配備される。ユーザ・サイトのホストは、仮想ボリューム・ストレージを使用することによりサービス・プロバイダ・サイトに配置されているサービス・プロバイダ・ストレージにアクセスできる。仮想ボリューム・ストレージを利用すると、ユーザはユーザ・サイトとサービス・プロバイダ・サイトの間で結合されているワイド・エリア・ネットワークに毎回アクセスする必要がなく便利である。

【００１３】本発明の一実施形態では、データ・ストレージ・サービスを提供するシステムであって、データ・ストレージ・サービスを提供するように構成されているサービス・プロバイダ・サイト及びサービス・プロバイダ・サイトにワイド・エリア・ネットワーク（ＷＡＮ）により結合されたユーザ・サイトを備え、ユーザ・サイトは仮想ストレージを持つローカル・ストレージを備え、仮想ストレージは同期ボリューム及び非同期ボリュームを備え、ローカル・ストレージは同期ボリューム内に書き込まれたデータをサービス・プロバイダ・サイトに即座に送信し、非同期ボリュームに書き込まれたデータを所定のスケジュールでサービス・プロバイダ・サイトに送信し、データがローカル・ストレージに格納されていない場合にサービス・プロバイダ・サイトからデータを読み込むように構成されているシステムを広範にわたって提示する。

【００１４】本発明の他の実施形態では、データの特徴に基づいてローカル・ストレージ内のデータを管理する方法を提示する。ローカル・ストレージ内の仮ボリューム・ストレージは、データが静的データかキャッシュされたデータか、同期データか非同期データかを判別する。仮想ボリューム・ストレージはさらに、ストレージ・ボリュームだけでなく、ディレクトリ、ファイル、シリンダ、及び／又はブロック・アドレスでもデータを管理できる。

【００１５】本発明の他の実施形態では、ユーザ・サイト内のユーザのアクセス活動を追跡する方法を提示する。ユーザ・アクセス活動を記録することで、サービス・プロバイダはユーザ・アクセス活動のパターン分析できる。この分析に基づき、サービス・プロバイダはユーザ・サイト内の仮想ボリューム・ストレージをチューニングできる。

【００１６】

【発明の実施の形態】以下では、当業者が本発明を実施し使用できるように説明し、また特定のアプリケーションとその要件の文脈において説明を行っている。実施形態のさまざまな修正は、当業者であればたやすく理解できるであろうし、またここで定義している一般的な原理は本発明の精神と範囲を逸脱することなく他の実施形態及びアプリケーションにも適用できる。そこで、本発明は示されている実施形態に限定されるのではなく、ここで開示されている原理、特徴および教示に一致する最も広い範囲を許容するものとする。

【００１７】本発明による一実施形態では、システム１００はユーザ・サイト１０１、さらに第１のサービス・プロバイダ・サイト１０２及び第２のサービス・プロバイダ・サイト１０３により構成される。ユーザ・サイト１０１の詳細を以下で説明する。システム１００内のサービス・プロバイダ・サイトの数は異なっていてもよい。第１のサービス・プロバイダ・サイト１０２は、ストレージ・ボリューム１２１、１２２、１２３、及び１２４を備え、第２のサービス・プロバイダ・サイト１０３はストレージ・ボリューム１２５、１２６、１２７、及び１２８を備える。サービス・プロバイダ・サイト内のストレージ・ボリュームの数は異なっていてもよい。サービス・プロバイダ・サイト１０２及び１０３は、たとえば、２つの異なる安全な建物内に配置できる。ユーザ・サイト１０１、第１のサービス・プロバイダ・サイト１０２、及び第２のサービス・プロバイダ・サイト１０３はそれぞれ、ワイド・エリア・ネットワーク（ＷＡＮ）１１０に接続されている。サービス・プロバイダ１０２及び１０３は、ＷＡＮ　１１０経由でデータ・ストレージ・サービスをユーザ・サイト１０１のユーザに提供する。

【００１８】当業者には周知のとおり、ＷＡＮとは、通常比較的広い範囲にわたる地理的領域にまたがるコンピュータ・ネットワークのことである。ＷＡＮはまた、ローカル・エリア・ネットワーク（ＬＡＮ）も含む。ＷＡＮに接続されているコンピュータは、多くの場合、電話システムなどの公衆網を通じて接続される。さらに、専用回線や衛星を通じて接続することもできる。現存最大のＷＡＮはインターネットである。

【００１９】ＷＡＮ　１１０では、各サイト（ユーザ・サイト１０１とサービス・プロバイダ・サイト１０２及び１０３）の間で通信ができる。各サイト１０１、１０２、及び１０３は、長い距離で互いに隔てられていてよい。ＷＡＮ　１１０は、通常、非同期転送モード（ＡＴＭ）、同期光ネットワーク（ＳＯＮＥＴ）、高密度波長

分割多重（ＤＷＤＭ）、又はインターネット・プロトコル（ＩＰ）ネットワークにより構成される。

【００２０】ユーザ・サイト１０１では、第１のホスト１０４はストレージ・エリア・ネットワーク（ＳＡＮ）１０７を介してローカル・ストレージ１０６に接続され、第２のホスト１０５はローカル・エリア・ネットワーク（ＬＡＮ）１０８を介してローカル・ストレージ１０６に接続されている。ホスト１０４及び１０５はローカル・ストレージ１０６にアクセスする。ホスト１０４及び１０５は、たとえば、サーバである。ユーザ・サイト１０１内のホストの数は異なっていてもよい。ＬＡＮ　１０８は、たとえば、ルータ１０９を介してＷＡＮ１１０に接続されている。本発明の教示から、ＳＡＮ１０７又はＬＡＮ　１０８のいずれか１つ、又はＳＡＮ　１０７及びＬＡＮ　１０８の両方を含むようにユーザ・サイト１０１を実装することができることを当業者は理解するであろう。

【００２１】当業者には周知のように、ＳＡＮは共有ストレージ・デバイスの高速サブネットワークであり、ＳＡＮによりＬＡＮ又はＷＡＮ内のすべてのサーバからすべてのストレージ・デバイスが利用可能である。ＳＡＮにさらにストレージ・デバイスを追加すると、その追加されたストレージ・デバイスも、さらに大きなネットワーク内の任意のサーバからアクセスできるようになる。

【００２２】ＳＡＮ　１０７は、たとえば、ファイバ・チャネル又は小型コンピュータ用周辺機器インタフェース（ＳＣＳＩ）で構成できる。当業者には周知のように、ファイバ・チャネルはコンピュータ及び大容量記憶デバイス製造業者のコンソーシアムが開発したシリアル・データ転送アーキテクチャであり、現在は、米国規格協会（ＡＮＳＩ）によって標準化されている。最も有名なファイバ・チャネル標準は、Fibre Channel Arbitrated Loop（ＦＣ－ＡＬ）で、新しい大容量記憶デバイスや非常に高い帯域幅を必要とするその他の周辺機器用に設計されている。光ファイバでデバイスを接続するＦＣ－ＡＬは、全二重データ転送を行い、転送速度は毎秒約１００メガバイト（ＭＢｐｓ）である。

【００２３】また当業者には周知のように、ＳＣＳＩは種々のコンピュータシステムで周辺装置をコンピュータに接続するために使用されているパラレル・インタフェース標準である。ＳＣＳＩインタフェースは、標準装備のシリアル・ポート及びパラレル・ポートに比べて高速なデータ転送速度（最大約８０ＭＢｐｓ）を利用できる。

【００２４】また当業者には周知のように、ＬＡＮとは、通常比較的狭い地域に置かれるコンピュータ・ネットワークのことである。ほとんどのＬＡＮは、単一の建物又は建物群に制限されている。ほとんどのＬＡＮは、ワークステーション及びパーソナル・コンピュータを接続する。ＬＡＮ内の各ノード（個々のコンピュータ）は、個々に中央処理装置（ＣＰＵ）を備え、プログラムを実行するが、ＬＡＮ上の任意の場所にあるデータ及びデバイスにアクセスすることも可能である。したがって、多くのユーザが、レーザ・プリンタなどの高価なデバイスをデータとともに共有できる。ユーザはさらに、ＬＡＮを使用して、たとえば電子メールを送信したりチャット・セッションに参加したりして互いに通信することもできる。ＬＡＮにはいろいろな種類のものがあるが、ＰＣにはイーサネット（登録商標）が最も一般的である。ＬＡＮは非常に高速にデータを送信することができ、電話回線でデータを送信するよりもかなり高速である。しかし、ＬＡＮの配線距離は制限されており、また単一のＬＡＮに接続できるコンピュータの台数にも制限がある。

【００２５】図１では、ＬＡＮ　１０８をたとえばイーサネット（登録商標）として構成している。第１のホスト１０４は、たとえば、ブロック（ＳＣＳＩ）インタフェース（Ｉ／Ｆ）を使用してローカル・ストレージ１０６にアクセスする。第２のホスト１０５は、たとえば、ファイル（ＮＦＳ又はＣＩＦＳ）Ｉ／Ｆを使用してローカル・ストレージ１０６にアクセスする。第２のホスト１０５は、たとえば、ブロック（ｉＳＣＳＩ）Ｉ／Ｆを使用してローカル・ストレージ１０６にアクセスすることもできる。

【００２６】ローカル・ストレージ１０６及びサービス・プロバイダ・サイト１０２及び１０３では、ＡＴＭ上でｉＳＣＳＩ又はファイバ・チャネル、ＳＯＮＥＴ上でファイバ・チャネル、又は独自ベンダ・プロトコルを使用できる。ローカル・ストレージ１０６は、仮想ストレージ１１１、コンフィギュレーション・テーブル１１２（図４も参照）、統計情報１１３（図５も参照）、アクセス・ログ１１４（図６も参照）、キャッシュ情報１１５、及びプログラム１１６を含む。コンポーネント１１２、１１３、１１４、１１５、及び１１６ではローカル・ストレージ１０６をキャッシュとして動作させることができる。ホスト１０４及び１０５は仮想ストレージ１１１にアクセスできる。仮想ストレージ１１１は、いくつかのボリューム１１７、１１８、１１９、及び１２０を含む。仮想ストレージ１１１内のボリュームの数は異なっていてもよい。これらのボリューム１１７～１２０の管理は、ローカル・ストレージ１０６が行う。ローカル・ストレージ１０６及びサービス・プロバイダ・サイト１０２及び１０３は、仮想ボリューム・ストレージ１１１の作成のため共同作業を行う。

【００２７】ユーザは、ユーザ・サイト１０１に仮想ボリューム・ストレージ１１１（ローカル・ストレージ１０６内の）を配備することができる。仮想ボリューム・ストレージ１１１がユーザ・サイト１０１に配備されると、ホスト１０４及び／又はホスト１０５のユーザは仮想ボリューム・ストレージ１１１を使用することによ

り、巨大なサービス・プロバイダ・ストレージ（ボリューム１２１～１２４及び／又はボリューム１２５～１２８）にアクセスできる。仮想ボリューム・ストレージ１１１は通常、サービス・プロバイダ・ストレージ・ボリューム（ボリューム１２１～１２４及び／又はボリューム１２５～１２８）に比べてサイズが小さい。仮想ボリューム・ストレージ１１１は、たとえば以下で説明するようなブロックＩ／Ｆ（ＳＣＳＩ）、ファイルＩ／Ｆ（ＮＦＳ、ＣＩＦＳ）などのいくつかのインタフェースを備える。仮想ボリューム・ストレージ１１１は、ユーザがＷＡＮ　１１０に毎回アクセスしなくてよいようにすると都合がよい。仮想ボリューム・ストレージ１１１ではさらに、サービス・プロバイダがユーザ向けに高速で信頼性の高いストレージ・システムを提供することもできる。

【００２８】図２は、ローカル・ストレージ１０６の一実施形態の詳細を示すブロック図である。ローカル・ストレージ１０６は、マイクロプロセッサ２０１（たとえば、Intel CorporationやMotorola Corporation製）、ローカル（内部）バス２０２、ＳＡＮ　１０７に結合されているＳＡＮインタフェース（Ｉ／Ｆ）２０３（図１）、ＬＡＮ　１０８に結合されているＬＡＮ　Ｉ／Ｆ　２０４（図１）、ＷＡＮ　１１０に結合されているＷＡＮ　Ｉ／Ｆ　２０５（図１）、メモリＩ／Ｆ２０６、メモリ２０７、ディスク・コントローラ２０８、及びローカル・ディスク２０９を備える。メモリ２０７は、コンフィギュレーション・テーブル１１２、統計情報１１３、アクセス・ログ１１４、キャッシュ情報１１５、及びプログラム１１６を格納する。マイクロプロセッサ２０１は、ローカル・ストレージ１０６内のすべてのリソースを制御し、プログラム１１６を使用してローカル・ストレージ１０６内のすべてのプロシージャを実行する。

【００２９】図２は、ローカル・ストレージ１０６内のプロシージャを実行しているときのローカル・ディスク２０９のスナップショットを示している。当業者には周知のとおり、スナップショットとは、実際のストレージの全ボリューム又はボリュームの一部のコピーである。図２に示されている情報及びプログラムはローカル・ディスク２０９に格納されている。ローカル・ストレージ１０６のブート手順で、これらのデータ及びプログラムはローカル・ディスク２０９からメモリ２０７に移動される。

【００３０】キャッシュ情報１１５は、キャッシュ・ディレクトリ２１０、保留データ・リスト２１１、及びキャッシュ・データ２１２で構成される。キャッシュ・ディレクトリ２１０は、メモリ２０７及びローカル・ディスク２０９内のキャッシュされたデータ２１２のディレクトリ（コンフィギュレーション情報）である。このキャッシュ・データ２１２は、ＬＲＵ（Least Recently Used）管理法で管理することができる。しかし、近い将来再びアクセスする可能性がないことで、ＬＲＵ管理の例外となるケースがある。このような例外の１つに順次アクセスがあり、ストレージ・アドレスが逐次的にアクセスされる。

【００３１】保留データ・リスト２１１は、ローカル・ストレージ１０６に保存されている保留データのリストである。保留データ・リスト２１１は、保留データへのポインタ、サービス・プロバイダ・サイトへのポインタ、及び同期期間などの各保留データの情報を含む。同期期間は、データの特徴により定義される。データが重要でない場合、このデータは所定の期間、ローカル・ストレージ１０６内に格納することができる。この期間は、たとえば、１分、１時間、１日、１週間、又は１カ月とすることができる。データが重要な場合、データが仮想ストレージ１１１に格納された後、コンフィギュレーション・テーブル１１２（図４）を使用してデータを即座に（同期して）サービス・プロバイダ・サイト（たとえば、サイト１０２）に格納すべきである。たとえば、同期データはボリューム００　１１７に格納される。図４で、ボリューム００　１１７は識別番号ＩＤ００で識別される。ボリューム００　１１７に格納されているデータは同期データなので、データはサービス・プロバイダ・サイト１０２（リモート・サイト１）のボリューム１０　１２１（リモートＩＤ　１０）に即座に格納される。

【００３２】この同期アクセス機能を使用すると、信頼性の高いデータ・ストレージを構築できる。ただし、ユーザ・サイト１０１からサービス・プロバイダ・サイト１０２及び／又は１０３にアクセスするのに時間がかかるため、同期アクセスにはかなりの時間が必要である。必要なアクセス時間は、システム１００で実行される特定のオペレーションによって異なる。

【００３３】ホスト１０４及び／又は１０５からアクセスされたデータはメモリ２０７に格納される。このデータは、キャッシュ・データ２１２と呼ばれる。キャッシュ・データ２１２は、ホスト１０４及び／又はホスト１０５のユーザによって頻繁にアクセスされる一時的データである。

【００３４】一実施形態では、プログラム１１６は、サーバ・プログラム２１３、シンクロナイザ２１７、キャッシュ制御２１８、スケジューラ２２０、及びデバイス・ドライバ２１９を含む。サーバ・プログラム２１３では、ローカル・ストレージ１０６及びホスト１０４（及び／又はホスト１０５）が互いに通信できる。サーバ・プログラム２１３は、ＮＦＳサーバ２１４、ＣＩＦＳサーバ２１５、及びＳＣＳＩサーバ２１６で構成される。ＮＦＳサーバ２１４は、従来のＮＦＳサーバとして動作する。ＣＩＦＳサーバ２１５、従来のＣＩＦＳサーバとして動作する。ＳＣＳＩサーバ２１６は、ＳＣＳＩター

ゲット・デバイスとして動作する。シンクロナイザ２１７は、保留データ・リスト２１１を使用して、ローカル・ストレージ１０６とサービス・プロバイダ・サイト１０２及び／又は１０３の同期をとる。キャッシュ制御２１８は、キャッシュ・ディレクトリ２１０を使用して、メモリ２０７とローカル・ディスク２０９内のキャッシュ・データ２１２を制御する。スケジューラ２２０は、サーバ・プログラム２１３、シンクロナイザ２１７、キャッシュ制御２１８、及びデバイス・ドライバ２１９などのすべてのプロシージャをスケジュールする。デバイス・ドライバ２１９は、ＳＡＮ　Ｉ／Ｆ　２０３、ＬＡＮ　Ｉ／Ｆ　２０４、ＷＡＮ　Ｉ／Ｆ　２０５、メモリ　Ｉ／Ｆ　２０６、及びディスク・コントローラ２０８などのローカル・ストレージ１０６内のすべてのデバイスを制御する。

【００３５】図３は、本発明の実施形態によるデータ・レイアウトの一例を示すブロック図である。ローカル・ストレージ１０６の管理は、以下でさらに詳しく説明しているように、データの特徴を使用することに基づいている。仮想ボリューム・ストレージ１１１では、データが静的かキャッシュされているか、同期か非同期かに応じてデータの特徴を取り扱う。以下でさらに詳しく説明しているように、仮想ボリューム・ストレージ１１１は、ストレージ・ボリュームだけでなく、ディレクトリ、ファイル、シリンダ、及び／又はブロック・アドレスでもデータの特徴を取り扱える。ユーザ・サイト１０１では、仮想ストレージ１１１に４種類の仮想ボリューム（ボリューム００　１１７、ボリューム０１　１１８、ボリューム０２　１１９、及びボリューム０３　１２０）がある。これらの仮想ボリュームの特徴である、（１）静的同期ボリューム、（２）静的非同期ボリューム、（３）キャッシュされた同期ボリューム、及び（４）キャッシュされた非同期ボリュームについて次に説明する。

【００３６】（１）静的同期ボリューム（ボリューム００　１１７）：ボリューム００　１１７は静的同期ボリュームである。したがって、ボリューム００　１１７は、第１のサービス・プロバイダ・サイト１０２において実際のストレージ・ボリュームであるボリューム１０　１２１と同じサイズを占有する。ホスト１０４及び／又は（ユーザ・サイト１０１の）は、ボリューム００　１１７からデータを読み込み、ボリューム１０　１２１からは読み込まない。アクセスされたデータがキャッシュ・データ・キャッシュ２１２（図２）内にある場合、ローカル・ストレージ１０６（図２）はローカル・ディスク２０９（図２）にアクセスせず、メモリ２０７にアクセスするだけである。ホスト１０４及び／又は１０５は、ボリューム００　１１７とボリューム１０　１２１にデータを書き込む。この静的非同期ボリューム、ボリューム００　１１７は、高速な読み込みアクセスと信頼性の高いストレージ・システムを実現する。

【００３７】（２）静的非同期ボリューム（ボリューム０１　１１８）：ボリューム０１　１１８は静的非同期ボリュームである。したがって、ボリューム００１　１１８は、第１のサービス・プロバイダ・サイト１０２において実際のストレージ・ボリュームであるボリューム１１　１２２と同じサイズを占有する。ホスト１０４及び／又は１０５は、ボリューム０１　１１８からだけデータを読み込み、ボリューム１１　１２２からはデータを読み込まない。アクセスされたデータがキャッシュ・データ２１２内にある場合、ローカル・ストレージ１０６はローカル・ディスク２０９にアクセスせず、メモリ２０７にアクセスするだけである。ホスト１０４及び／又は１０５は、ボリューム０１　１１８にデータを書き込み、保留データ・リスト２１１（図２）に登録する。この保留データは、以下で説明するように、所定のスケジュールで同期処理される。

【００３８】この保留データは、バックグラウンド・ジョブでローカル・ストレージ１０６（図２）内に圧縮し、ＷＡＮ　１１０のトラフィックを低減することができる。この静的非同期ボリューム、ボリューム０１　１１８は、ボリュームの高速な読み書きを行うが、信頼度は同期ボリュームの場合と同じではないことがある。

【００３９】（３）キャッシュされた同期ボリューム（ボリューム０２　１１９）：ボリューム０２　１１９はキャッシュされた同期ボリュームである。したがって、ボリューム０２　１１９は、第１のサービス・プロバイダ・サイト１０２において実際のストレージ・ボリュームであるボリューム１２　１２３と同じサイズを占有する。ボリューム０２　１１９はキャッシュとして動作する。したがって、ホスト１０４及び／又は１０５によってアクセスされるデータがボリューム０２　１１９内にない。アクセスされたデータがローカル・ディスク２０９（図２）内にない場合、ローカル・ストレージ１０６（図２）は第１のサービス・プロバイダ・サイト１０２からデータを読み込んで、読み込んだデータをローカル・ディスク２０９に書き込む。データがローカル・ストレージ１０６内に入った後、ローカル・ストレージ１０６はメモリ２０７を介してローカル・ディスク２０９からデータをホスト１０４及び／又は１０５に送る。アクセスされたデータがキャッシュ・データ２１２内にある場合、ローカル・ストレージ１０６はローカル・ディスク２０９にアクセスせず、メモリ２０７にアクセスするだけである。ホスト１０４及び／又は１０５は、ボリューム０２　１１９とボリューム１２１２３にデータを書き込む。このキャッシュされた同期ボリューム、ボリューム０２　１１９は都合よく、ローカル・ストレージ１０６のサイズを減らすことができる。この仮想ボリューム、ボリューム０２　１１９ではさらに、信頼性の高いストレージ・システムを構築できる。

【００４０】（４）キャッシュされた非同期ボリューム（ボリューム０３　１２０）：ボリューム０３　１２０はキャッシュされた非同期ボリュームである。したがって、ボリューム０３　１２０は、第１のサービス・プロバイダ・サイト１０２において実際のストレージ・ボリュームであるボリューム１３　１２４と同じサイズを占有する。ボリューム０４　１２０はキャッシュとして動作する。したがって、ホスト１０４及び／又は１０５によってアクセスされるデータがボリューム０３　１２０内にない。アクセスされたデータがローカル・ディスク２０９内にない場合、ローカル・ストレージ１０６は第１のサービス・プロバイダ・サイト１０２からデータを読み込んで、この読み込んだデータをローカル・ディスク２０９に書き込む。データがすでにローカル・ストレージ１０６内に入っていた場合その後、ローカル・ストレージ１０６はメモリ２０７を介してローカル・ディスク２０９からデータをホスト１０４及び／又は１０５に送る。アクセスされたデータがデータ・キャッシュ２０７内にある場合、ローカル・ストレージ１０６はローカル・ディスク２０９にアクセスせず、メモリ２０７にアクセスするだけである。ホスト１０４及び／又は１０５は、ボリューム０１　１１８にデータを書き込み、保留データ・リスト２１１（図２）に登録する。この保留データは、あるスケジュールで同期処理される。このキャッシュされた非同期ボリューム、ボリューム０３　１２０は都合よく、ローカル・ストレージ１０６のサイズを減らすことができるが、信頼性は同期ボリュームと同じにならない場合がある。

【００４１】図３はさらに、サービス・プロバイダ・サイトで提供しているサービスも示している。第１のサービス・プロバイダ・サイト１０２のボリューム１１　１２２は、第２のサービス・プロバイダ・サイト１０３内のボリューム２１　１２６に複製される。ボリューム１２　１２３は、第２のサービス・プロバイダ・サイト１０３内のボリューム２２　１２７に複製される。ボリューム１３　１２４は、第２のサービス・プロバイダ・サイト１０３内のボリューム２３　１２８に複製される。ボリューム・データのこの複製（リモート・コピー１３０）は、障害回復方法となる。ボリューム・データは、たとえば、従来のリモート・ミラーリング技術を使用して複製できる。さらに、第１のサービス・プロバイダ・サイト１０２では、ユーザ・サイト１０１の機能がなくてもボリューム１２　１２３のデータのバックアップを実行できる（矢印１３２を参照）。このバックアップ・サービスにより、ユーザはデータ・バックアップの作業負荷を軽減できる。

【００４２】以下で説明しているように、仮想ボリューム・ストレージ１１１では、ホスト１０４及び／又は１０５からのアクセス活動に基づいてトレース・データを作成できる。仮想ボリューム・ストレージ１１１は、サービス・プロバイダに対して、ユーザ・アクセス・パターンの分析のためこのトレース・データを用意することができる。ユーザ・アクセス・パターンを分析した後、サービス・プロバイダは仮想ボリューム・ストレージ１１１を都合よくチューニングできる。

【００４３】図４は、本発明の実施形態によるコンフィギュレーション・テーブル１１２の一例を示すブロック図である。コンフィギュレーション・テーブル１１２は、パラメータ「ＩＤ」、「インタフェース」、「サブエリア」、「リモート・サイト」、「リモートＩＤ」、「ボリューム・タイプ」、「サービス」（「バックアップ」、「分析」）、「サイズ」（「合計」、「使用済み」、「空き」）、及び「ポリシー」で構成される。「ＩＤ」は、ローカル・ストレージ１０６内の仮想ボリュームのボリューム識別を示す。たとえば、ボリューム００　１１７（図３）のＩＤは００である。「インタフェース」は、ＳＣＳＩ、ＮＦＳ、及びＣＩＦＳなどの仮想ボリューム・インタフェース・タイプを示す。「サブエリア」は、仮想ボリューム内のサブエリアの個数を示す。サブエリアは、ブロック・アドレス、シリンダ、ファイル、及び／又はディレクトリのあるエリアとすることができる。「ボリューム・タイプ」は、サブエリア間で異なることがある。この場合、図４の実例で、各サブエリアはボリューム・タイプが同じである。たとえば、ボリューム００　１１７（図３）（ＩＤは００）は上述のように同じ静的同期ボリューム・タイプである。「リモート・サイト」は、リモート・サイトの識別子を示す（たとえば、サービス・プロバイダ・サイト１０２の識別子）。これは、ｈｔｔｐアドレスのＩＰアドレスでもよい。「リモートＩＤ」は、サービス・プロバイダ・サイト内のボリュームＩＤを示す。たとえば、サービス・プロバイダ・サイト１０２のボリューム１０　１２１はリモートＩＤ　１０を持つ。「ボリューム・タイプ」は、静的な又はキャッシュされた、同期又は非同期などのボリュームのタイプを示す。「サービス」は、バックアップ・サービスや分析サービスなどのサービス・プロバイダが提供するサービスを示す。サービス・プロバイダは、分析サービスを提供するときに、統計情報１１３（図２）へのポインタを設定し、統計情報（アクセス・ログ）１１４を作成し、データの特徴を取得する。分析サービスが提供される場合、図４の例に示されているように「Ｐｔｒ」がコンフィギュレーション・テーブル１１２の分析セクションで表示される。たとえば、コンフィギュレーション・テーブル１１２のボリューム０２　１１９（ＩＤ０２）は、分析セクションにポインタＰｔｒを持つ（行４００を参照）。（ボリューム０２　１１９の）ＩＤ　０２に対するこのポインタは、図５内の統計情報１１３の中のＩＤ　０２を指している。ＩＤ　０２は、サブエリア（ディレクトリ）「/usra」（行５００を参照）を持ち、ポインタ（「ptr」）はログＰｔｒ

セクション内にある。このポインタ（ptr）は、図６の例のボリューム０２　１１９に対し作成されたアクセス・ログであるアクセス・ログ１１４を指している。

【００４４】さらに、図5の例では、（ボリューム０２　１１９の）ＩＤ　０２はディレクトリ「/usrb」（行５０５参照）を含み、「Ｎｕ１１」値（ポインタなし）がログＰｔｒセクション内にある。さらに、ＩＤ　０２は、ディレクトリ「/usrc」（行５１０を参照）を含み、Ｎｕ１１値（ポインタなし）がログＰｔｒセクション内にある。したがって、ディレクトリ/usrb及び/usrcにアクセス・ログを指すポインタがないため、これらのディレクトリに対し関連するアクセス・ログ１１４が作成されていない。

【００４５】コンフィギュレーション・テーブル１１２で、「サイズ」は、ギガバイト単位で合計メモリ・サイズ（「合計」）を示し、また使用済みサイズ（「使用済み」）及び空きサイズ（「空き」）を示している。ボリュームがファイル・システムのローカル・ストレージ１０６内にない場合、合計サイズのみ判明している必要がある。パラメータ「ポリシー」は、同期のスケジュールを示す。たとえば、行４１０では、ボリューム００　１１７（ＩＤは００）は、図４で「ポリシー」パラメータに「Ｎｕ１１」値が設定されているが、このＮｕ１１値は同期スケジュールがボリューム００　１１７に対し設定されていないことを示す。ボリューム０２　１１９（ＩＤは０２）は、「ポリシー」パラメータで「週」値を設定しており、この値は、ボリューム０２　１１９内のデータが週に１度サービス・プロバイダ・サイト（たとえば、サイト１０２）のデータと同期処理されていることを示す。ボリューム０３　１２０（ＩＤは０３）は、「ポリシー」パラメータで「日」値を設定しており、この値は、ボリューム０３　１２０内のデータが毎日サービス・プロバイダ・サイト（たとえば、サイト１０２）のデータと同期処理されていることを示す。

【００４６】コンフィギュレーション・テーブル１１２のパラメータは、コンフィギュレーション・テーブル１１２内のコンフィギュレーション・データを設定することにより設定される。各ストレージ・サブシステムは関連するコンフィギュレーション・テーブルを持つ。

【００４７】図5は、本発明の実施形態による統計情報１１３の一例を示すブロック図である。ユーザ又はサービス・プロバイダがコンフィギュレーション・テーブル１１２（図４）で「サービス分析」パラメータを「Ｐｔｒ」に設定した場合、次のことが可能になる。ローカル・ストレージ１０６は、統計情報１１３のこのテーブルでローカル・ストレージ１０６のユーザ統計アクセス情報を収集する。「統計情報」は、「ＩＤ」、「サブエリア」、「ＩＯ／ｓ」及び「ＭＢ／ｓ」（「平均、最大読み込み」、「平均、最大書き込み」）、「ヒット率」、及び「ログＰｔｒ」で構成される。パラメータ「ＩＤ」及び「サブエリア」（図５）は、コンフィギュレーション・テーブル１１２（図４）の「ＩＤ」及び「サブエリア」と同じ意味を持つ。「ＩＯ／ｓ」は、１秒あたりの読み込み及び書き込みコマンドなど、１秒あたりのホスト／ユーザ入力又は出力活動の個数を示す。ボリューム又はボリューム・ディレクトリ内の読み込み又は書き込みアクセスは、知られている適当な方法で記録できる。

【００４８】「ＭＢ／ｓ」は１秒あたりのバイト数を示す。ローカル・ストレージ１０６は、ボリューム（又はボリューム内のサブエリア）での読み込み及び書き込みの発生数を収集し、この発生の平均（Ａｖｅ）をとる。ローカル・ストレージ１０６は、さらに、ボリューム又はボリューム内のサブエリアでの読み込み及び書き込みの発生の最大発生回数（Ｍａｘ）も追跡する。

【００４９】「ヒット率」は、ローカル・ストレージのヒット率を示す。このヒット率は、たとえば、ローカル・ディスク２０９（図２）だけのものである。ヒット率は、次の式（１）で定義される。

（１）ヒット率＝１００％（#ローカル・ディスク読み込み／#全読み込み）

ただし、「#ローカル・ディスク読み込み」は、ホスト別のローカル・ディスク２０９内の読み込み回数であり、「#全読み込み」は、ホスト別の合計読み込み回数である。

【００５０】統計情報１１３は、サービス・プロバイダ・サイト１０２及び／又は１０３に定期的に送られる。サービス・プロバイダでは、データの特徴を分析し、この分析結果からサービス・プロバイダはユーザ・サイト１０１のホストのユーザに適切な解決策を提示することができる。

【００５１】上述のように、ログＰｔｒは、アクセス・ログ１１４へのポインタを示している。サービス・プロバイダは、ユーザ・アクセス・パターンの詳細を知りたい場合、このポインタを作成されたアクセス・ログ１１４に設定する。

【００５２】図6は、本発明の実施形態によるアクセス・ログ１１４の一例を示すブロック図である。サービス・プロバイダは、ユーザのアクセス・パターンの詳細を分析するときに、アクセス・ログ１１４を作成し、ログＰｔｒ（図５の統計情報１１３内にある）のポインタをこのアクセス・ログ１１４に設定する。ローカル・ストレージ１０６は、ユーザ別に各アクセスのアクセス・ログ１１４を収集する。アクセス・ログ１１４は、パラメータ「日付」、「時刻」、「コマンド」、「ファイルＩＤ」、「アドレス」、及び「サイズ」で構成される。

「日付」は、ユーザがアクセスした日付を示す。「時刻」は、アクセスの時刻を示す。「コマンド」は、アクセスのコマンド・タイプ（たとえば、読み込み又は書き込み）を示す。「ファイルＩＤ」は、このコマンドでアクセスしたファイルの識別を示す。アクセス・ログがＳ

ＣＳＩ用であれば、「ファイルＩＤ」はＮｕｌｌとすることもできる。「アドレス」及び「サイズ」は、アクセス・アドレス及びサイズを示す。

【００５３】読み込みプロセス

読み込みプロシージャでは、キャッシュされたデバイス（ローカル・ストレージ１０６）はキャッシュとして動作する。ホスト１０４及び／又は１０５は、ローカル・ストレージ１０６にアクセスするが、これは、ローカル・ストレージ１０６の空き容量が大きいからである。まず、ホスト（たとえば、ホスト１０４又はホスト１０５）は、ＬＡＮ　１０８（図１）を介してＮＦＳ、ＣＩＦＳ、又はｉＳＣＳＩプロトコルの読み込みコマンドを発行するか、又はＳＡＮ　１０７を介してＳＣＳＩプロトコルの読み込みコマンドを発行する。ローカル・ストレージ１０６は、ＬＡＮ　Ｉ／Ｆ　２０４又はＳＡＮ　Ｉ／Ｆ　２０３（図２）を介して読み込みコマンドを受信する。デバイス・ドライバ２１９（図２）は、ホストからのこのコマンドを処理し、このコマンドをスケジューラ２２０に入れる。次にスケジューラ２２０は読み込みコマンドを分析して、読み込みコマンドをサーバ・プログラム２１３内において適切なサーバ（ＮＦＳサーバ２１４、ＣＩＦＳサーバ２１５、又はＳＣＳＩサーバ２１６）に入れて読み込みコマンドを処理することができる。各サーバ・プログラム２１３は、読み込みコマンドで要求されたデータがキャッシュ・データ２１２（図２）内にあるかどうかを調べる。すべてのデータ（読み込みコマンドで要求された）がキャッシュ・データ２１２内にある場合は、ローカル・ストレージ１０６は内部バス２０２とネットワーク・インタフェース（ＳＡＮ　Ｉ／Ｆ　２０３又はＬＡＮ　Ｉ／Ｆ　２０４）を介してデータを要求側ホスト（ホスト１０４又は１０５）に返す。他方、データ（読み込みコマンドで要求された）の全部又は一部がキャッシュ・データ２１２内にない場合、要求されたデータはローカル・ディスク２０９から、又はサービス・プロバイダ・サイト１０２又は１０３からキャッシュ・データ２１２に移動すべきである。

【００５４】図７は、本発明の実施形態によるこの読み込みプロセスの流れ図である。ローカル・ストレージ１０６がホスト（たとえば、ホスト１０４又は１０５）から読み込みコマンドを受け取った後、ローカル・ストレージ１０６はデータのボリューム・タイプを調べる（ステップ７０１）。スケジューラ２２０は、ボリュームＩＤをチェックすることによりコンフィギュレーション・テーブル１１２内のボリューム・タイプ（たとえば、静的タイプ）を調べる。特に、スケジューラ２２０（図２）は、読み込みコマンドを分析して、データのボリューム・タイプを判別し、また読み込みコマンドを処理する（取り扱う）サーバ・プログラム（サーバ２１４、２１５、又は２１６）を決定する。スケジューラ２２０は、コンフィギュレーション・テーブル１１２（図４）を見て読み込みコマンドを処理するためのボリューム・タイプと適切なサーバ・プログラムを調べる。ボリューム・タイプが静的ボリュームの場合、ローカル・ストレージ１０６はステップ７０２、７０３、及び７０４を飛ばし、以下で説明するステップ７０５に進む。ボリューム・タイプがキャッシュされたボリュームの場合、ローカル・ストレージ１０６はデータ（読み込みコマンドで要求した）がキャッシュ・データ２１２内に格納されているかいないかを調べる（ステップ７０２）。

【００５５】ローカル・ストレージ１０６は、キャッシュ・ディレクトリ２１０を調べる（ステップ７０３）。特に、サーバ・プログラム２１３の適切なサーバ（２１４、２１５、又は２１６）は、読み込みコマンドで要求したデータがローカル・ディスク２０９内にあるかどうかを調べる。すべてのデータ（読み込みコマンドで要求した）がローカル・ディスク２０９に格納されている場合、ローカル・ストレージ１０６はステップ７０４を飛ばして、以下で説明するステップ７０５に進む。データ（読み込みコマンドで要求した）がローカル・ディスク２０９内にない場合、ローカル・ストレージ１０６はステップ７０４を実行する。

【００５６】ステップ７０４で、ローカル・ストレージ１０６はローカル・ディスク２０９内でデータ領域を割り当て、サービス・プロバイダ・サイト（たとえば、サービス・プロバイダ・サイト１０２）からデータ（読み込みコマンドで要求した）を読み込む。特に、キャッシュ制御２１８では、コンフィギュレーション・テーブル１１２（図４）を使用してデータが取り出されるサービス・プロバイダ・サイトを決定する。キャッシュ制御２１８では、サービス・プロバイダ・サイトからデータを読み込み、データをローカル・ディスク２０９に格納する。

【００５７】データをローカル・ディスク２０９に格納した後、ローカル・ストレージ１０６はローカル・ディスク２０９からキャッシュ・データ２１２へデータを移動する（書き込む）（ステップ７０５）。特に、キャッシュ制御２１８は、ローカル・ディスク２０９からキャッシュ・データ２１２にデータを移動する。続いてローカル・ストレージ１０６は、統計情報１１３（図５）を更新する（ステップ７０６）。特に、ローカル・ストレージ１０６のスケジューラ２２０は、統計情報１１３を更新する。統計情報１１３内のログ・ポインタ（ログＰｔｒ）（図５）が設定されている場合、ローカル・ストレージ１０６はこの読み込みコマンドのログ・データをアクセス・ログ１１４（図６）に追加する。スケジューラ２２０又はデバイス・ドライバ２１９は、統計情報１１３とアクセス・ログ１１４を作成する。通常、統計情報１１３とアクセス・ログ１１４の作成においてはスケジューラ２２０が用いられる。

【００５８】その後、データ（読み込みコマンドで要求

した）はキャッシュ・データ２１２から、読み込みコマンドを送ったホスト（たとえばホスト１０４又は１０５）に移動される。特に、キャッシュ制御２１８はキャッシュ・データ２１２からデータを適切なサーバ・プログラム（２１４、２１５、又は２１６）に移動し、適切なサーバ・プログラム（２１４、２１４、又は２１６）がデータを、読み込みコマンドを送ったホストに返す。

【００５９】書き込みプロセス

書き込みプロシージャでは、キャッシュされたデバイス（ローカル・ストレージ１０６）はキャッシュとして動作する。ホスト１０４及び／又は１０５は、ローカル・ストレージ１０６にアクセスするが、これは、ローカル・ストレージ１０６の空き容量が大きいからである。まず、ホスト（たとえば、ホスト１０４又はホスト１０５）は、ＬＡＮ　１０８を介してＮＦＳ、ＣＩＦＳ、又はｉＳＣＳＩプロトコルの書き込みコマンドを発行するか、又はＳＡＮ　１０７を介してＳＣＳＩプロトコルの書き込みコマンドを発行する。ローカル・ストレージ１０６は、ＬＡＮ　Ｉ／Ｆ　２０４又はＳＡＮ　Ｉ／Ｆ　２０３を介して書き込みコマンドを受信する。デバイス・ドライバ２１９（図２）は、ホストからのこの読み込みコマンドを処理し、このコマンドをスケジューラ２２０に入れる。次にスケジューラ２２０は書き込みコマンドを分析して、書き込みコマンドをサーバ・プログラム２１３内において適切なサーバ（ＮＦＳサーバ２１４、ＣＩＦＳサーバ２１５、又はＳＣＳＩサーバ２１６）に入れる。各サーバ・プログラム２１３は、書き込みコマンドのデータがキャッシュ・データ２１２（図２）内にあるかどうかを調べる。すべてのデータ（書き込みコマンドの）がキャッシュ・データ２１２内にある場合は、ローカル・ストレージ１０６は内部バス２０２とネットワーク・インタフェース（ＳＡＮ　Ｉ／Ｆ　２０３又はＬＡＮ　Ｉ／Ｆ　２０４）を介してホストからデータを受信する。ローカル・ストレージ１０６は、ローカル・ディスク２０９及びキャッシュ２１２内の割り当てられた空き領域にデータを上書きする。他方、（書き込みコマンドの）データの全部又は一部がキャッシュ・データ２１２内にない場合、ローカル・ストレージ１０６は、キャッシュ・データ２１２内にないデータの残りについてある領域（ローカル・ディスク２０９及びキャッシュ・データ２１２内の）を割り当てる。すべてのデータ領域の割り当てが済んだら、ローカル・ストレージ１０６はその割り当てられた領域内にデータを格納する。すべてのデータをキャッシュ・データ２１２内に格納する場合、ローカル・ストレージ１０６はローカル・ディスク２０９にデータを格納し、このデータをプロバイダ・サイト（たとえば、サービス・プロバイダ・サイト１０２）に送信する。

【００６０】図８は、本発明の実施形態による書き込みプロセスの流れ図である。ローカル・ストレージ１０６がホストから書き込みコマンドを受け取った後、ローカル・ストレージ１０６は書き込みコマンドのデータのボリューム・タイプを調べる（ステップ８０１）。特に、スケジューラ２２０（図２）は、書き込みコマンドを分析して、データのボリューム・タイプを判別し、また読み込みコマンドを処理する（取り扱う）サーバ・プログラム（サーバ２１４、２１５、又は２１６）を決定する。スケジューラ２２０は、コンフィギュレーション・テーブル１１２（図４）を見て書き込みコマンドを処理するためのボリューム・タイプと適切なサーバ・プログラムを調べる。ボリューム・タイプが静的ボリュームの場合、ローカル・ストレージ１０６はステップ８０２、８０３、及び８０４を飛ばし、以下で説明するステップ８０５に進む。ボリューム・タイプがキャッシュされたボリュームの場合、ローカル・ストレージ１０６はデータ（書き込みコマンドの）がキャッシュ・データ２１２（図２）内に格納されているかいないかを調べる。ローカル・ストレージ１０６は、キャッシュ・ディレクトリ２１０（図２）を調べる（ステップ８０２）。

【００６１】特に、サーバ・プログラム２１３の適切なサーバ（２１４、２１５、又は２１６）は、書き込みコマンドのデータがローカル・ディスク２０９内にあるかどうかを調べる。ステップ８０３で、すべてのデータ（書き込みコマンドで要求した）がローカル・ディスク２０９内で割り当てられている場合、ローカル・ストレージ１０６はステップ８０４を飛ばして、ステップ８０５に進む。ステップ８０３で、データがローカル・ディスク２０９内にない場合、ローカル・ストレージ１０６はステップ８０４を実行する。

【００６２】（書き込みコマンドの）データがローカル・ディスク２０９内で割り当てられていない場合、ローカル・ストレージ１０６は書き込みデータについてローカル・ディスク２０９とキャッシュ・データ２１２の両方にデータ領域を割り当てる（ステップ８０４）。特に、キャッシュ制御２１８は、ローカル・ディスク２０９及びキャッシュ・データ２１２の両方でデータを割り当てる。

【００６３】ローカル・ストレージ１０６は、キャッシュ・データ２１２を介して書き込みデータをローカル・ディスク２０９に書き込む。特に、適切なサーバ・プログラム（ＮＦＳサーバ２１４、ＣＩＦＳサーバ２１５、又はＳＣＳＩサーバ２１６）はデータをローカル・ディスク２０９に書き込む。

【００６４】ローカル・ストレージ１０６は、コンフィギュレーション・テーブル１１２（図４）を使用してデータを書き込む際のボリューム・タイプを調べる（ステップ８０６）。特に、サーバ・プログラム（ＮＦＳサーバ２１４、ＣＩＦＳサーバ２１５、又はＳＣＳＩサーバ２１６）はコンフィギュレーション・テーブル１１２を使用してボリューム・タ

イプが同期領域の場合、ローカル・ストレージ１０６は即座にサービス・プロバイダ・サイト（たとえば、サイト１０２）とデータの同期をとり（書き込み）（ステップ８０８）、以下で説明するステップ８０９に進む。シンクロナイザ２１７（図２）はサービス・プロバイダ・サイトとの（ユーザ・サイト内の）データの同期処理を実行する。ローカル・ストレージ１０６内のキャッシュ制御２１８は、コンフィギュレーション・テーブル１１２を使用してサービス・プロバイダ・サイトに格納すべきデータを認識する。

【００６５】ステップ８０６で、ボリューム・タイプが非同期領域の場合、ローカル・ストレージ１０６は保留データ・リスト２１１（図２）を更新する（ステップ８０７）。特に、サーバ・プログラム（サーバ２１４、２１５、又は２１６）が保留データ・リストを更新する。

【００６６】続いてローカル・ストレージ１０６は、統計情報１１３（図５）を更新する（ステップ８０９）。特に、スケジューラ２２０は、統計情報１１３を更新する。統計情報１１３内のログ・ポインタ（ログＰｔｒ）（図５）が設定されている場合、ローカル・ストレージ１０６はこの書き込みコマンドのログ・データをアクセス・ログ１１４に追加する。特に、スケジューラ２２０は書き込みコマンドのログ・データをアクセス・ログ１１４に追加する。これで書き込みプロセスは終了する。

【００６７】同期プロセス

非同期ボリューム書き込み（非同期領域へのデータ書き込み）の場合、データは、ユーザ又はサービス・プロバイダによって定められたスケジュールに従ってサービス・プロバイダ・サイト（たとえば、サイト１０２）に送信される。このスケジュールは、たとえば、図４のコンフィギュレーション・テーブル１１２内の「ポリシー」エントリ内の値により定められる。スケジューラ２２０（図２）は、シンクロナイザ２１７（図２）を定期的に実行することで、サービス・プロバイダ・サイトとデータの同期をとることができる。この期間は、システム１００設定に応じて、たとえば、約１．０ミリ秒又は１０．０ミリ秒に設定できる。図９は、本発明の実施形態による同期プロセスの流れ図である。同期プロセスは、シンクロナイザ２１７（図２）によって実行できる。

【００６８】シンクロナイザ２１７は、保留データ・リスト２１１（図２）内のヘッド・データを選択する（ステップ９０１）。最初に、シンクロナイザ２１７は、保留データ・リスト２１１を調べる（ステップ９０２）。保留データ・リスト２１１にデータがなければ、シンクロナイザ２１７は同期プロセスを終了する。同期すべきデータが１つ又は複数（保留データ・リスト２１１内に）ある場合、シンクロナイザ２１７は以下で説明するようにステップ９０２～９０５を実行する。

【００６９】シンクロナイザ２１７は、保留データ・リスト２１１の情報を調べる（ステップ９０３）。保留データ・リスト２１１内のこの保留データの同期をとる必要があれば、シンクロナイザ２１７は保留データをサービス・プロバイダ・サイト（たとえば、サイト１０２）に送り、保留データとサービス・プロバイダ・サイトとの同期をとる（ステップ９０４）。ローカル・ストレージ１０６内のキャッシュ制御２１８は、コンフィギュレーション・テーブル１１２（図４）内の「リモートＩＤ」及び「リモート・サイト」の値に基づいてサービス・プロバイダ・サイト（たとえば、サイト１０２）に格納すべきデータを認識する。

【００７０】シンクロナイザ２１７は、保留データ・リスト２１１内の次のデータを選択する（ステップ９０５）。シンクロナイザ２１７は、保留データ・リスト２１１内で選択するデータがなくなるまでステップ９０２～９０５を繰り返す。保留データ・リスト２１１に選択するデータがなくなったら、図９の方法は終了する。コンピュータが上述のいずれかの方法を実行できるようにコンピュータで読み取り可能な媒体に格納できるプログラム又はコードを実装することも本発明の範囲である。

【００７１】そこで、本発明について特定の実施形態を引用しながら説明してきたが、前記開示においては修正、さまざまな変更、及び置き換えを許容することが意図されており、規定されているとおり本発明の範囲を逸脱することなく他の特徴の対応する使用がなくても場合によっては本発明のいくつかの特徴を採用することは認められるであろう。

【図面の簡単な説明】

【図１】本発明の実施形態によるシステムのブロック図。

【図２】図１に示されているローカル・ストレージの一実施形態の詳細を示すブロック図。

【図３】本発明の実施形態によるデータ・レイアウトの一例を示すブロック図。

【図４】本発明の実施形態によるコンフィギュレーション・テーブルの一例を示すブロック図。

【図５】本発明の実施形態による統計情報の一例のブロック図。

【図６】本発明の実施形態によるアクセス・ログの一例のブロック図。

【図７】本発明の実施形態による読み込みプロセスの流れ図。

【図８】本発明の実施形態による書き込みプロセスの流れ図。

【図９】本発明の実施形態による同期プロセスの流れ図。

【符号の説明】

１０１…ユーザ・サイト、１０２，１０３…サービス・プロバイダ・サイト、１０４，１０５…ホスト、１０６…ローカル・ストレージ、１０７…ＳＡＮ、１０８…ＬＡＮ、１１１…仮想ストレージ

【図１】

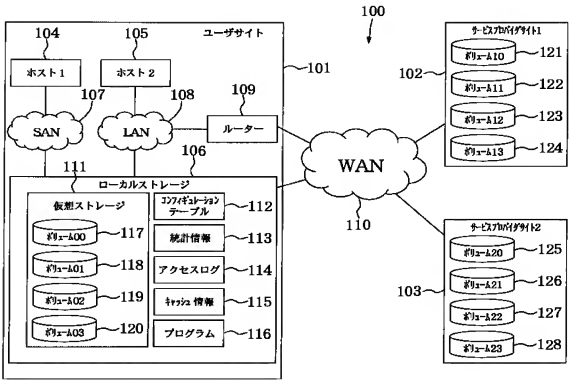
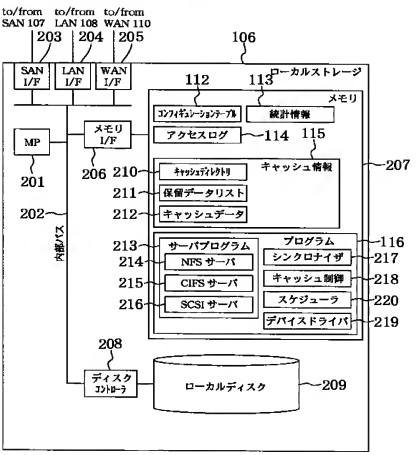


図 1

【図２】

図 2



【図６】

図 6

114
アクセスログ

日付	時刻	コマンド	ファイル ID	アドレス	サイズ
2000.8.1	0:0:0	読み込み	0	0x0000	0x0001
2000.8.1	0:0:0	書き込み	0	0x0000	0x0001
2000.8.1	0:0:1	読み込み	10	0x0100	0x0100
2000.8.1	0:0:1	読み込み	10	0x0200	0x0100
2000.8.1	0:0:2	読み込み	10	0x0200	0x0100
2000.8.1	0:0:2	読み込み	10	0x0300	0x0100
2000.8.1	0:0:2	読み込み	10	0x0400	0x0100

【図３】

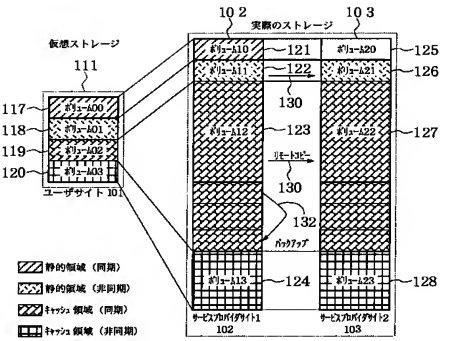


図３

【図４】

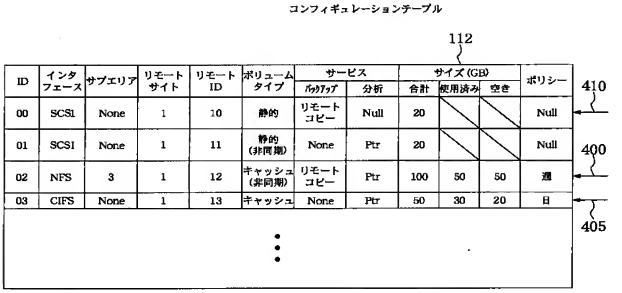


図４

【図５】

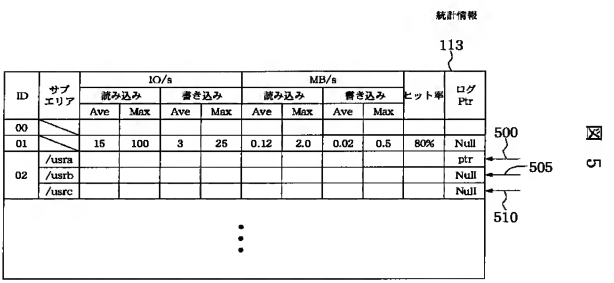
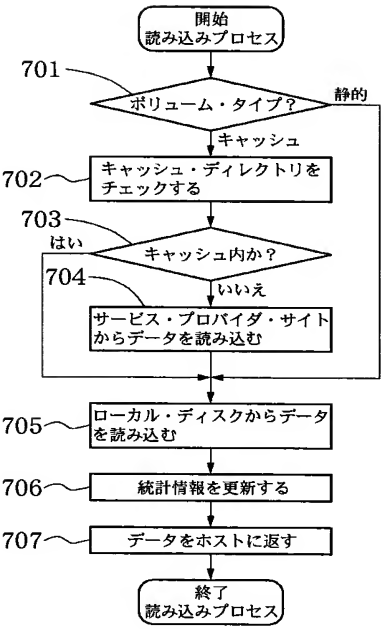


図５

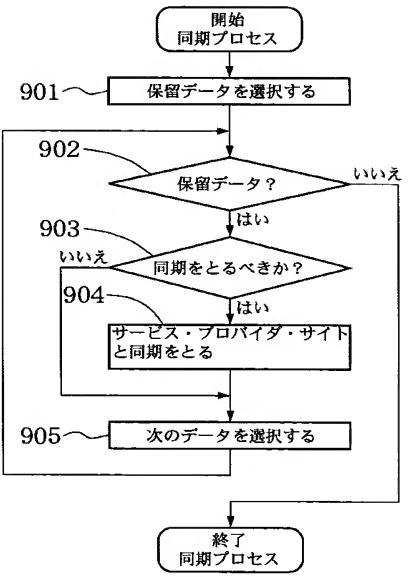
【図7】

図　7


開始
読み込みプロセス
701
ボリューム・タイプ？
静的
キャッシュ
702
キャッシュ・ディレクトリを
チェックする
703
はい
キャッシュ内か？
704
いいえ
サービス・プロバイダ・サイト
からデータを読み込む
705
ローカル・ディスクからデータ
を読み込む
706
統計情報を更新する
707
データをホストに返す
終了
読み込みプロセス

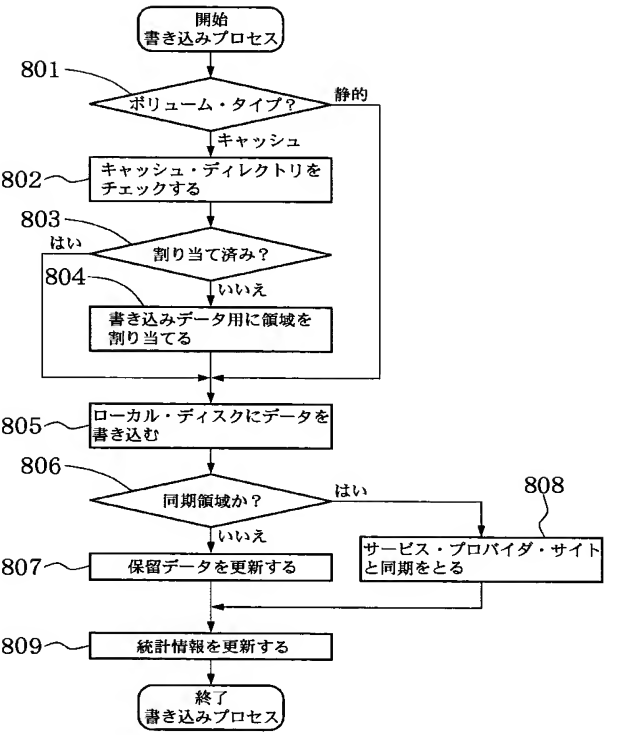
【図9】

図　9


開始
同期プロセス
901
保留データを選択する
902
保留データ？
いいえ
はい
903
いいえ
同期をとるべきか？
904
はい
サービス・プロバイダ・サイト
と同期をとる
905
次のデータを選択する
終了
同期プロセス

【図８】

図　8



フロントページの続き

Ｆターム(参考)　5B065 CE21 EA35 EK05
5B082 FA11 FA12 GB06 HA02

【外国語明細書】

1 Title of Invention

VIRTUAL VOLUME STORAGE

2 Claims

1. A system for providing a data storage service, the system comprising:

a service provider site configured to provide a data storage service; and

a user site coupled by a wide area network (WAN) to the service provider site, the user site comprising a local storage having a virtual storage, the virtual storage having a synchronous volume and an asynchronous volume, the local storage configured to immediately transmit to the service provider site data that is written in the synchronous volume, to transmit at a predetermined schedule to the service provider site data that is written in the asynchronous volume, and to read data from the service provider site if the data is not stored in the local storage.

2. The system of claim 1 wherein the local storage is configured to record an access activity to the local storage by the host.

3. The system of claim 1 wherein the local storage records an access activity to a volume in the virtual storage.

4. The system of claim 1 wherein the local storage records an access activity in a sub area of a volume in the virtual storage.

5. The system of claim 2 wherein the access activity is provided to the service provider site for analysis.

6. A method of providing a data storage service, the method comprising:

providing a local storage having a virtual storage, the virtual storage comprising a synchronous volume and an asynchronous volume; and

selectively performing at least one of the following:

immediately transmitting to a service provider site data that is written in the synchronous volume of the virtual storage;

transmitting at a predetermined schedule to the service provider site data that is written in the asynchronous volume the virtual storage; and

reading data from the service provider site if the data is not stored in the local storage.

7. The method of claim 6, further comprising:

recording an access activity to the local storage.

8. The method of claim 7, further comprising:

providing a record of the access activity to the service provider site.

9. An electronically-readable medium storing a program for permitting a computer to perform a method of providing a data storage service, the method comprising:

selectively performing at least one of the following:

immediately transmitting to a service provider site data that is written in a synchronous volume of a virtual storage;

transmitting at a predetermined schedule to the service provider site data that is written in an asynchronous volume the virtual storage; and

reading data from the service provider site if the data is not stored in a local storage that includes the virtual storage.

10. A program code embedded on a carrier wave for causing a computer to perform a method of providing a data storage service, the method comprising:

selectively performing at least one of the following:

immediately transmitting to a service provider site data that is written in a synchronous volume of a virtual storage;

transmitting at a predetermined schedule to the service provider site data that is written in an asynchronous volume the virtual storage ; and

reading data from the service provider site if the data is not s tored in a local storage that includes the virtual storage.

11. An apparatus for processing data in a data storage system, the appar atus comprising:

a user site comprising a local storage having a virtnal storage, the virtual storage including a synchronous area and an asynchronous area, the local storage configured to immediately transmit to a service provid er site data that is written to the synchronous area, to transmit at a p redetermined schedule to the service provider site data that is written to the asynchronous area, and to read data from the service provider sit e if the data is not stored in the local storage.

12. A method of processing data in a data storage system, the method com prising:

in response to a command signal to a local storage, selectively perf orming at least one of the following:

immediately transmitting to a service provider site data that is written to a synchronous area in a virtual storage;

transmitting at a predetermined schedule to the service provider site data that is written to an asynchronous area in the virtual storag e;

reading data from the service provider site if the data is not s tored in the local storage.

13. The method of reading data in a data storage system, the method comp rising:

receiving a read command from a host to a local storage;

determining the volume type of the data that is requested by the rea

d command;

if the volume type is static, then reading the data from a local dis k in the local storage;

if the volume type is cached, then checking if the data is stored in the local disk;

if the data is stored in the local disk, then reading the data from the local disk;

if the data is not stored in the local disk, then reading the data f rom a remote service provider site, storing the data in the local disk, and reading the data from the local disk;

after reading the data from the local disk, updating statistical inf ormation relating to the read command: and

returning the data to the host.

14. An apparatus for processing data in a data storage system, the appar atus comprising:

a service provider site configured to immediately receive data that is written to a synchronous area in a virtual storage, to receive at a p redetermined schedule data that is written to an asynchronous area in th e virtual storage, and to transmit data to a local storage that includes the virtual storage if the data is not stored in the local storage.

15. A method of processing data in a data storage system, the method com prising:

selectively performing one of the following:

immediately receiving data that is written to a synchronous area in a virtual storage;

receiving at a predetermined schedule data that is written to an asynchronous area in the virtual storage; and

transmitting data to a local storage that includes the virtual s torage if the data is not stored in the local storage.

16. A method of writing data in a data storage system, the method compri sing:

receiving a write command from a host to a local storage;

determining the volume type of the data of the write command;

if the volume type is static, then writing the data to a local disk in the local storage;

if the volume type is cached, then checking if the data is stored in the local disk;

if the data is stored in the local disk, then writing the data to th e local disk from a cache;

if the data is not stored in the local disk, then allocating a data area in the local disk and in the cache and then writing the data to the local disk from the cache;

checking if the volume type of the data is synchronous;

if the volume type is synchronous, then immediately synchronizing th e date with a remote service provider site and then updating statistical information relating to the write command; and

if the volume type is not synchronous, then synchronizing the data w ith a remote service provider site based on a predetermined schedule and then updating statistical information relating to the write command.

17. A system for providing a data storage service, the system comprising :

means for selectively performing at least one of the following:

immediately transmitting to a service provider site data that is written in a synchronous volume of a virtual storage;

transmitting at a predetermined schedule to the service provider site data that is written in an asynchronous volume the virtual storage ; and

reading data from the service provider site if the data is not s

tored in a local storage that iocludes the virtual storage.

3 Detailed Description of Invention

FIELD OF THE INVENTION

The present invention relates to data storage systems, and relates m ore particularly to a system and method for providing a virtual volume d ata storage in a user site.

BACKGROUND OF THE INVENTION

Managing a large data storage system is very difficult. Typically, a data storage service provider provides a storage-related service such as providing storage volumes, data backup, and the like. From a user si te, if a user connects to this storage via the storage service provider' s directory, then the user must use a long distance connection to connec t to this storage. This process causes more delays than if a user is co nnecting to a local storage.

The World Wide Web (WWW) effectively acts as a storage system in a w ide area. The user may deploy a proxy server in the user site. As know n to those skilled in the art, a proxy server is a server that sits betw een a client application (such as a web browser) and a remote server. T he proxy server provides a cache of items that are available on the remo te servers. The proxy server intercepts all requests that are made to t he remote server so that the proxy server can determine if it can instea d fulfill the request. If the proxy server is unable to fulfill the req uest, then the proxy server will forward the request to the remote serve r for processing. A proxy server can just only read cache data from the remote server or site, and it does not support a write procedure to the remote server or site. Additionally, the proxy server can not provide a service that depends on the data feature. As a result, the proxy serv

er causes an ineffective usage of resources, and the use of a proxy server also causes more delays than the use of a local storage.

U.S. patent 5,155,835 discloses a multilevel, hierarchical, dynamically mapped data storage subsystem. This patent reference does not disclose storage systems between wide area networks (WANs) and ignores data features such as access interface type (block or file, and the like), purpose (user data or system file or backup file and the like). Thus, the subsystem disclosed in this reference is inefficient.

The Internet protocol small computer system interface (iSCSI) uses the Internet Protocol (IP) networking infrastructure to quickly transport large amounts of block storage (SCSI) data over existing local area and/or wide area networks. With the potential to support all major networking protocols, iSCSI (IP SAN) can unify network architecture across an entire enterprise, thereby reducing the overall network cost and complexity. To ensure reliability, iSCSI can use known network management tools and utilities that have been developed for IP networks. The iSCSI protocol is discussed, for example, at the website, http://www.ece.cmu.edu/~ips/index.html) in the IP Storage section, http://www.ece.cmu.edu/~ips/Docs/docs.html.

The working group of Internet Engineering Task Force (IETF) at http://www.ietf.org provides a network file system (NFS) version 3 (rfc1813 of IETF) and a common Internet File system (CIFS) protocol (http://www.cifs.org).

The NFS is an open operating system that allows all network users to access shared files that are stored in different types of computers. NFS provides access to shared files through an interface called Virtual File System (VFS) which runs on top of the Transmission Control Protocol/Internet Protocol (TCP/IP). With NFS, computers connected to a network can operate as clients while accessing remote files and as servers while

providing remote users access to local shared files.

The CIFS protocol defines a standard for remote file access using millions of computers at a time. With CIFS, users with different platforms and computers can share files without having to install new software. CIFS runs over TCP/IP, but uses the Server Message Block (SMB) protocol found in Microsoft Windows for file and printer access. Therefore, CIFS will allow all applications (including Web browsers) to open and share files across the Internet.

The websites http://www.cnt.com and http://www.san.com describe wide area network (WAN) connections to a storage or storage area networks (SANs). The WAN connection may be an asynchronous transfer mode (ATM), synchronous optical network (SONET), and the like.

The above references do not disclose methods for providing hierarchical management techniques between storage systems and host devices.

There is a need for a system and method that will overcome the above-mentioned deficiencies of conventional methods and systems. There is also a need for a system and method that will permit an access service provider to have a reliable storage system and that will permit a user to quickly access the storage system. There is also a need for a system and method that will permit an access service provider to be able to tune a local storage system based upon user statistic data and user log data.

SUMMARY

The present invention may advantageously provide a reliable virtual local storage in a user site. The present invention may also advantageously provide a virtual local storage that permits faster access for a user. The present invention may also advantageously permit a service provider to tune up the local storage by using user statistics data and user log data that are tracked by the local storage, and the service provide

r may then charge the user for these tuning services.

In one embodiment, the present invention permits a service provider to provide a data storage service via a wide area network. A virtual volume storage is deployed at the user site. The virtual volume storage allows a host(s) at the user site to access the service provider storage located at the service provider site. The virtual volume storage advantageously permits the user to avoid having to access each time the wide area network coupled between the user site and the service provider site.

In one embodiment, the present invention broadly provides a system for providing a data storage service, comprising: a service provider site configured to provide a data storage service; and a user site coupled by a wide area network (WAN) to the service provider site, the user site comprising a local storage having a virtual storage, the virtual storage having a synchronous volume and an asynchronous volume, the local storage configured to immediately transmit to the service provider site data that is written in the synchronous volume, to transmit at a predetermined schedule to the service provider site data that is written in the asynchronous volume, and to read data from the service provider site if the data is not stored in the local storage.

In another embodiment, the present invention provides a method of managing data in a local storage based on the data feature. The virtual volume storage in the local storage determines if the data is static or cached, and synchronous or asynchronous. The virtual volume storage can also manage data not only by storage volume, but also by directory, file, cylinder, and/or block address.

In another embodiment, the present invention provides a method of tracing the access activities of a user in the user site. The user access activities are recorded to permit the service provider to analyze patterns in the user access activities. Based on this analysis, the service

provider can tune the virtual volume storage in the user site.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The following description is provided to enable any person skilled in the art to make and use the present invention, and is provided in the context of a particular application and its requirements. Various modifications to the embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the present invention. Thus, the present invention is not intended to be limited to the embodiments shown, but is to be accorded the widest scope consistent with the principles, features, and teachings disclosed herein.

In one embodiment in accordance with the present invention, a system 100 is configured by a user site 101, and by a first service provider site 102 and a second service provider site 103. The details of the user site 101 is described below. The number of service provider sites in the system 100 may vary. The first service provider site 102 includes storage volumes 121, 122, 123, and 124, while the second service provider site 103 includes storage volumes 125, 126, 127, and 128. The number of storage volumes in a service provider site may vary. The service provider sites 102 and 103 may be located in, for example, two different safe buildings. The user site 101, first service provider site 102, and second service provider site 103 are each connected to a Wide Area Network (WAN) 110. The service providers 102 and 103 provide data storage services via the WAN 110 to a user at the user site 101.

As known to those skilled in the art, a WAN is a computer network that typically spans a relatively large geographical area. A WAN may also

include local area networks (LANs). Computers connected to a WAN are often connected through public networks, such as the telephone system. They can also be connected through leased lines or satellites. The largest WAN in existence is the Internet.

The WAN 110 permits communication between each site (user site 101 and service provider sites 102 and 103). Each site 101, 102, and 103 may be separated from each other by long distances. The WAN 110 is typically configured by asynchronous transfer mode (ATM), synchronous optical network (SONET), Dense Wavelength Division Multiplexing (DWDM), or Internet Protocol (IP) network.

At the user site 101, a first host 104 is connected to a local storage 106 via a storage area network (SAN) 107, while a second host 105 is connected to the local storage 106 via a local area network (LAN) 108. The hosts 104 and 105 access the local storage 106. The hosts 104 and 105 are, for example, servers. The number of hosts in the user site 101 may vary. The LAN 108 is connected to the WAN 110 via, for example, a router 109. From the teachings of the present invention herein, it is understood by those skilled in the art that the user site 101 may be implemented to include only one of the SAN 107 or LAN 108, or both the SAN 107 and LAN 108.

As known to those skilled in the art, a SAN is a high-speed sub-network of shared storage devices wherein the SAN makes all storage devices available to all servers in a LAN or WAN. As more storage devices are added to a SAN, these added storage devices will also be accessible from any server in the larger network.

The SAN 107 may be configured with, for example, fibre channel or Small Computer System Interface (SCSI). As known to those skilled in the art, a fibre channel is a serial data transfer architecture developed by a consortium of computer and mass storage device manufacturers and now

being standardized by the American National Standards Institute (ANSI). The most prominent fibre channel standard is the Fibre Channel Arbitrated Loop (FC-AL) which is designed for new mass storage devices and other peripheral devices that require a very high bandwidth. Using optical fibers to connect the devices, FC-AL supports full-duplex data transfer rates of approximately 100 megabytes per second (MBps).

As also known to those skilled in the art, SCSI is a parallel interface standard used by Apple Macintosh computers, personal computers (PCs), and many UNIX systems for attaching peripheral devices to computers. SCSI interfaces provide for faster data transmission rates (up to about 80 MBps) than standard serial and parallel ports.

As also known to those skilled in the art, a LAN is a computer network that typically spans a relatively small area. Most LANs are confined to a single building or group of buildings. Most LANs connect workstations and personal computers. Each node (individual computer) in a LAN has its own central processing unit (CPU) with which it executes programs, but it is also able to access data and devices anywhere on the LAN. Thus, many users can share expensive devices, such as laser printers, as well as data. Users can also use the LAN to communicate with each other, by, for example, sending e-mail or engaging in chat sessions. There are many different types of LANs, with Ethernets being the most common for PCs. LANs are capable of transmitting data at very fast rates, much faster than the data transmitted over a telephone line. However, the distances over LANs are limited, and there is also a limit on the number of computers that can be attached to a single LAN.

In Figure 1, the LAN 108 is configured as, for example, an Ethernet.

The first host 104 accesses the local storage 106 by use of, for example, a block (SCSI) interface (I/F). The second host 105 accesses the local storage 106 by, for example, a file (NFS or CIFS) I/F. The second

host 105 may also access the local storage 106 by, for example, a block (iSCSI) I/F.

The local storage 106 and service provider sites 102 and 103 may use iSCSI or fiber channel over ATM, fiber channel over SONET, or a unique vendor protocol.

The local storage 106 includes a virtual storage 111, configuration table 112 (see also Figure 4), statistical information 113 (see also Figure 5), access log 114 (see also Figure 6), cache information 115, and programs 116. The components 112, 113, 114, 115, and 116 permit the local storage 106 to act like a cache. The hosts 104 and 105 can access the virtual storage 111. The virtual storage 111 includes some volumes 117, 118, 119 and 120. The number of volumes in the virtual storage 111 may vary. These volumes 117-120 are managed by the local storage 106. The local storage 106 and service provider sites 102 and 103 collaborate for creation of the virtual volume storage 111.

The user can deploy the virtual volume storage 111 (in local storage 106) at the user site 101. The virtual volume storage 111 allows the users of host 104 and/or host 105 to access the huge service provider storage (volumes 121-124 and/or volumes 125-128) as the virtual volume storage 111 is deployed at the user site 101. The virtual volume storage 111 typically has a smaller size than the service provider storage volumes (volumes 121-124 and/or volumes 125-128). The virtual volume storage 111 has several interfaces such as, for example, a block I/F (SCSI), file I/F (NFS, CIFS) as discussed below. The virtual volume storage 111 advantageously permits the user to avoid in having to access the WAN 110 every time. The virtual volume storage 111 also enables the service provider to provide a faster and more reliable storage system for the user.

Figure 2 is a block diagram showing the details of one embodiment of

the local storage 106. The local storage 106 includes a microprocessor 201 (which is available from, for example, Intel Corporation or Motorola Corporation), a local (internal) hns 202, a SAN interface (I/F) 203 coupled to the SAN 107 (Figure 1), a LAN I/F 204 coupled to the LAN 108 (Figure 1), a WAN I/F 205 coupled to the WAN 110 (Figure 1), a memory I/F 206, a memory 207 , a disk controller 208, and a local disk 209. The memory 207 contains the configuration table 112, statistical information 113, access log 114, cache information 115, and programs 116. The microprocessor 201 controls all resources in the local storage 106 and executes all procedures in the local storage 106 by using the programs 116.

Figure 2 shows a snapshot of the local disk 209 during the running of procedures in the local storage 106. As known to those skilled in the art, a snapshot is a copy of a whole volume or a part of a volume of a real storage. These information and programs shown in Figure 2 are stored in the local disk 209. At the boot sequence of local storage 106, these data and programs are moved from the local disk 209 to the memory 207.

The cache information 115 is configured with cache directory 210, pending data list 211, and cache data 212. The cache directory 210 is a directory (configuration information) of the cached data 212 in memory 207 and local disk 209. This cache data 212 may be managed by the least recently used (LRU) management. But some cases should be an exception of the LRU management, because there will be no possibility to access again in the near future. One such exception is sequential access where the storage address is accessed in a sequential manner.

The pending data list 211 is a list of pending data which is saved in the local storage 106. The pending data list 211 has information of each pending data such as a pointer to pending data, a pointer to a service provider site, and a synchronous period. A synchronous period is def

ined by the data feature. If the data is not important, then this data may be stored in the local storage 106 for a predetermined period. This period may he, for example, one minute, one hour, one day, one week, or one month. If the data is important, then after the data is stored in the virtual storage 111. the data should be stored immediately (synchronous) in a service provider sitc (e.g., site 102) hy using the configuration table 112 (Figure 4). For example, synchronous data is stored in Volume 00 117. In Figure 4, Volnmc 00 117 is identified with the identification number ID 00. Since the data stored in Volume 00 117 is synchronous data, the data will be immediately stored in the service provider site1 102 (Remote Site 1) at Volume 10 121 (Remote ID 10).

This synchronous access feature provides a reliable data storage system. However, synchronous access requires much time because of the access time by tbe user site 101 to the service provider sites 102 and/or 103. The required access time depends on the particular operation being carried out on the system 100.

Data that are accessed by hosts 104 and/or 105 are stored in the memory 207. This data is called cache data 212. Cache data 212 is a temporary data that is frequently accessed hy a user of host 104 and/or host 105.

In one embodiment, the programs 116 include server programs 213, a synchronizer 217, a cache control 218, a scheduler 220, and a device driver 219. The server programs 213 permit the local storage 106 and the host 104 (and/or host 105) to communicate with each other., The server programs 213 are configured with an NFS server 214, a CIFS server 215, and an SCSI server 216. The NFS server 214 acts as a conventional NFS server. The CIFS server 215 acts as a conventional CIFS server. The SCSI server 216 acts as a SCSI target device. The synchronizer 217 synchronizes the local storage 106 and the service provider sites 102 and/or 103 hy

using the pending data list 211. The cache control 218 controls cache 212 in the memory 207 and local disk 209 by using the cache directory 210. The scheduler 220 schedules all procedures such as the server programs 213, synchronizer 217, cache control 218, and device driver 219. The device driver 219 controls all devices in local storage 106 such as the SAN I/F 203, the LAN I/F 204, the WAN I/F 205, the memory I/F 206, and the disk controller 208.

Figure 3 is a block diagram showing one example of a data layout in accordance with an embodiment of the present invention. The management of the local storage 106 is based upon the use of data feature, as described below in additional details. The virtual volume storage 111 deals with the data feature, depending on whether the data is static or cached, synchronous or asynchronous. As also described further below, the virtual volume storage 111 can deal with data feature not only by storage volumes, but also by directory, file, cylinder, and/or block address. At user site 101, there are four (4) types of virtual volumes (Volume 00 117, Volume 01 118, Volume 02 119, and Volume 03 120) in the virtual storage 111. The features of these virtual volumes are described bellow: (1) the static synchronous volume, (2) the static asynchronous volume, (3) the cached synchronous volume, and (4) the cached asynchronous volume.

(1) The static synchronous volume (Volume 00 117):

Volume 00 117 is a static synchronous volume. Thus, Volume 00 117 occupies the same size as the real storage volume, Volume 10 121, in the first service provider site 102. The hosts 104 and/or 105 (at user site 101) read the data from Volume 00 117 and not from Volume 10 121. If the accessed data is in the cache data cache 212 (Figure 2), then the local storage 106 (Figure 2) does not access the local disk 209 (Figure 2)

and just accesses the memory 207. The hosts 104 and/or 105 write data t o both Volume 00 117 and Volume 10 121. This static synchronous volume, Volume 00 117, provides a fast read access and reliable storage system.

(2) The static asynchronous Volume (Volume 01 118):

Volume 01 118 is a static asynchronous volume. Thus, Volume 001 118 occupies the same size as the real storage volume, Volume 11 122, in th e first service provider site 102. The hosts 104 and/or 105 read data fr om just only Volume 01 118 and not from Volume 11 122. If the accessed data is in the cache data 212, then local storage 106 does not access th e local disk 209 and only accesses the memory 207. The hosts 104 and/or 105 write data to Volume 01 118 and register to the pending data list 2 11 (Figure 2). This pending data is synchronized with a predetermined s chedule, as described below.

This pending data may be compressed in the background job in the loc al storage 106 (Figure 2) to reduce the WAN 110 traffic. This static as ynchronous volume, Volume 01 118, provides a fast read and write volume, but may not provide the same reliability as a synchronous volume.

(3) The cached synchronous volume (Volume 02 119):

Volume 02 119 is a cached synchronous volume. Thus, Volume 02 119 o ccupies same size as the real storage volume, Volume 12 123, in the firs t service provider site 102. Volume 02 119 acts as a cache. Thus, ther e is no data in Volume 02 119 that is accessed by the hosts 104 and/or 1 05. If the data is not in the local disk 209 (Figure 2), then the local storage 106 (Figure 2)reads data from the first service provider site 1 02 and writes the read data to the local disk 209. After the data is in the local storage 106, the local storage 106 sends the data from the lo cal disk 209 via memory 207 to the hosts 104 and/or 105. If the accesse

d data is in the cache data 212, then the local storage 106 does not acc ess local disk 209 and just accesses the memory 207. The hosts 104 and/ or 105 write data to Volume 02 119 and Volume 12 123. This cached synch ronous volume, Volume 02 119, advantageously permits the reduction of si ze of the local storage 106. This virtual volume, Volume 02 119, also p rovides a reliable storage system.

(4) The cached asynchronous volume (Volume 03 120):

Volume 03 120 is a cached asynchronous volume. Thus, Volume 03 120 occupies same size as the real storage volume, Volume 13 124, in the fir st service provider site 102. Volume 04 120 acts as a cache. Thus ther e is no data in Volume 03 120 that is accessed by the hosts 104 and/or 1 05. If the data is not in the local disk 209, the local storage 106 rea ds data from the first service provider site 102 and writes this read da ta into the local disk 209. After the data is already in the local stor age 106, the local storage 106 sends the data from local disk 209 via me mory 207 to the hosts 104 and/or 105. If the accessed data is in the da ta cache 207, then the local storage 106 does not access local disk 209 and just accesses the memory 207. The hosts 104 and/or 105 write the da ta to Volume 01 118 and register to pending data list 211 (Figure 2). T his pending data is synchronized with a schedule. This cached asynchron ous volume, Volume 03 120, advantageously permits the reduction of size of the local storage 106, but may not provide the same reliability as a synchronous volume.

Figure 3 also illustrates a service provided by the service provider site. Volume 11 122 in the first service provider site 102 is duplicat ed on volume 21 126 in the second service provider site 103. Volume 12 123 is duplicated on volume 22 127 in the second service provider site 1 03. Volume 13 124 is duplicated on volume 23 128 in the second service

provider site 103. This duplication (remote copying 130) of volume data provides a disaster recovery method. The volume data may be duplicated by use of, for example, conventional remote mirroring technology. Additionally, at the first service provider site 102, a backup of the data in Volume 12 123 may be made (see arrow 132) without requiring the function of the user site 101. This backup service permits the user to reduce the workload of data back up.

As now discussed below, the virtual volume storage 111 can make a trace data based upon access activities from the hosts 104 and/or 105. The virtual volume storage 111 can provide to the service provider this trace data for purposes of analyzing the user access pattern. After analyzing the user access pattern, the service provider can advantageously tune the virtual volume storage 111.

Figure 4 is a block diagram of an example of a configuration table 112 in accordance with an embodiment of the present invention. The configuration table 112 is configured with the following parameters: "ID", "Interface", "Sub Area", "Remote Site", "Remote ID", "Volume Type", "Service" ("Backup", "Analyze"), "size" ("Total", "Used", "Free"), and "Policy". ID indicates volume identification of a virtual volume in the local storage 106. For example, Volume 00 117 (Figure 3) has an ID 00. Interface indicates the virtual volume interface type such as SCSI, NFS, and CIFS.. Sub area indicates the number of sub areas in the virtual volume. A sub area may be a certain area of block address, cylinders, file, and/or directory. Volume type may be different between each sub area. In this case in the example of Figure 4, each sub area has the same volume type. For example, Volume 00 117 (Figure 3) (with ID 00) is a static synchronous volume type as mentioned previously. Remote Site indicates an identifier of the remote site (e.g., the identifier of service provider

site 102). This may be the IP address of an http address. Remote ID indicates volume ID in a service provider site. For example, Volume 10 121 in service provider site 102 has a remote ID 10. Volume type indicates a type of volume such as static or cached, synchronous or asynchronous. Service indicates a service provided by the service provider Such as a back up service or an analyze service. When a service provider provides an analyze service, the service provider sets a pointer to a statistical information 113 (Figure 2) and creates statistical information (access log) 114 to obtain a data feature. If an analyze service is provided, then "Ptr" will be indicated in the Analyze section in the Configuration Table 112 as shown in the example of Figure 4. For example, Volume 02 119 (ID 02) in Configuration Table 112 has a pointer Ptr in the Analyze section (see row 400). This pointer for ID 02 (of Volume 02 119) points to ID 02 in the Statistical Information 113 in Figure 5. ID 02 has a Sub area (directory) "/usra" (see row 500) with a pointer ("ptr") in the Log Ptr section. This pointer (ptr) points to the access log 114 which is an access log created for Volume 02 119 in the example of Figure 6.

It is further noted that in the example of Figure 5, the ID 02 (of Volume 02 119) includes a directory "/usrb" (see row 505) with a "Null" value (no pointer) in the Log Ptr section. Additionally, ID 02 includes a directory "/usrc" (see row 510) with a Null value (no pointer) in the Log Ptr section. Thus, no associated access logs 114 have been created for the directories /usrb and /usrc since these directories do not have pointers that point to an access log.

In the Configuration Table 112, Size indicates total memory size (Total) in gigabytes, used size (Used), and free size (Free). If a volume is not in the file system in local storage 106, then only the total size needs to be known. The parameter policy indicates a schedule for synchronization. For example, in row 410, the Volume 00 117 (with ID 00) has

a "null" value set in the Policy parameter in Figurer 4, and this null value indicates that a synchronization schedule has not been set for Volume 00 117. Volume 02 119 (with ID 02) has a "week" value set in the Policy parameter, and this value indicates that the data in Volume 02 119 is synchronized once per week with data in a service provider site (e.g., site 102). Volume 03 120 (with ID 03) has a "day" value set in the Policy parameter, and this value indicates that the data in Volume 03 120 is synchronized daily with data in a service provider site (e.g., site 102).

The parameters in the Configuration Table 112 are set by setting the configuration data in the Configuration Table 112. Each storage subsystem has an associated Configuration Table.

Figure 5 is a block diagram showing an example of statistical information 113 in accordance with an embodiment of the present invention. When a user or service provider sets the Service-Analyze parameter to "Ptr" in the Configuration Table 112 (Figure 4), then the following will be enabled. The local storage 106 collects the user statistical access information of local storage 106 in this table of Statistical Information 113. The Statistical Information is configured with "ID", "Sub Area", "IO/s" and "MB/s" (Read Ave, Max, Write Ave, Max), "Hit Ratio", and "Log Ptr". The parameters ID and Sub area (in Figure 5) have the same meanings as ID and Sub area in Configuration Table 112 (Figure 4). IO/s indicates the number of host/user input or output activities per second such as read and write commands per second. A read or write access in a volume or in a volume directory may be recorded by use of any suitable known methods.

MB/s indicates number of bytes per second. The local storage 106 collects each read and write occurrence in a volume (or in a sub area in a

volume), and averages (Ave) this occurrences. The local storage 106 al so tracks the maximum occurrence (Max) of read and write occurrence in a volume or in a sub area in a volume.

Hit Ratio indicates a hit ratio of local storage. This hit ratio is , for example, only just for the local disk 209 (Figure 2). A hit ratio is defined in equation (1):

(1) Hit ratio = 100%(#LOCAL DISK READ/#ALL READS)

where #LOCAL DISK READ is the number of reads in the local disk 209 by a host, and #ALL READS is the number of total reads by the host.

Statistical Information 113 is sent to the service provider sites 10 2 and/or 103 periodically. The service provider analyzes the feature of data, and from this analysis the service provider can propose better so lutions to the user of a host at the user site 101.

As stated above, Log Ptr indicates a pointer to the access log 114. If the service provider wants to know more detail of user access patter ns, then the service provider can set this pointer to a created accesses log 114.

Figure 6 is a block diagram showing an example of an access log 114 in accordance with an embodiment of the present invention. When the ser vice provider wants to analyze the details of a user's access pattern, t hen the service provider creates an access log 114 and sets a pointer at Log Ptr (in Statistical Information 113 in Figure 5) to this access log 114. The local storage 106 collects an access log 114 of each access by a user. This access log 114 is configured with the parameters: "Date", "Time", "Command", "File ID", "Address", and "Size". Date indicates th e date of an access by a user. Time indicates the time of an access. C ommand indicates the command type (e.g., read or write) of an access. F ile ID indicates the identification of a file that was accessed by this

command. If access log is for SCSI, the File ID may be null. Address a nd size indicates the access address and size.

Read Process

In the read procedure, the cached device (local storage 106) acts as cache. The hosts 104 and/or 105 access the local storage 106, since th e local storage 106 has a large space. At first, a host (e.g., host 104 or host 105) issues a read command of NFS, CIFS, or iSCSI protocol via LAN 108 (Figure 1), or a read command of SCSI protocol via SAN 107. The local storage 106 receives the read command via LAN I/F 204 or SAN I/F 203 (Figure 2). The device driver 219 (Figure 2) handles this read comma nd from a host and places this read command into the scheduler 220. The scheduler 220 then analyzes the read command and places the read comman d into a proper server (NFS server 214, CIFS server 215, or SCSI server 216) within the server programs 213 for purposes of processing the read command. Each server program 213 checks whether the data requested by t he read command is in or not in the cache data 212 (Figure 2). If all d ata (which is requested by the read command) is in the cache data 212, t hen the local storage 106 returns data to the requesting host (host 104 or 105) via internal bus 202 and a network interface (SAN I/F 203 or LAN I/F 204). On the other hand, if all of or part of the data (requested by the read command) is not in cache data 212, then the requested data s hould be moved to cache data 212 from the local disk 209 or from the ser vice provider sites 102 or 103.

Figure 7 is a flowchart diagram of this read process in accordance w ith an embodiment of the present invention. After the local storage 106 receives a read command from a host (e.g., host 104 or 105), the local storage 106 checks 701 for the volume type of the data. The scheduler 2 20 checks configuration table 112 for the volume type (e.g., static type

) by check a volume ID. In particular, the scheduler 220 (Figure 2) ana lyzes the read command to determine the volume type of the data and to d etermine which server program (server 214, 215, or 216) should process (handle) the read command. The scheduler 220 looks at configuration tabl e 112 (Figure 4) for the volume type and the appropriate server program to handle the read command. If volume type is a static volume, then the local storage 106 skips the steps 702, 703, and 704 and proceeds to ste p 705 which is discussed below. If the volume type is a cached volume, then the local storage 106 checks 702 if the data (requested by the read command) is stored in or not stored in the cache data 212.

The local storage 106 checks 703 the cache directory 210. In partic ular, the appropriate server (214, 215, or 216) in the server program 21 3 checks whether the data requested by the read command is in or not in the local disk 209. If all data (requested by the read command) is sto red in the local disk 209, then the local storage 106 skips step 704 and proceeds to step 705 which is described below. If data (request by the read command) is not in the local disk 209, then local storage 106 exec utes step 704.

In step 704, the local storage 106 allocates data area in the local disk 209 and reads data (requested by the read command) from a service p rovider site (e.g., service provider site 102). In particular, the cach e control 218 uses the configuration table 112 (Figure 4) to determine t he service provider site from where data should be obtained. The cache control 218 reads the data from the service provider site and stores the data in the local disk 209.

After data is stored in local disk 209, the local storage 106 will m ove (read) 705 the data from the local disk 209 to the cache data 212. In particular, the cache control 218 moves the data from the local disk 209 to the cache data 212. The local storage 106 then updates 706 the s

tatistical information 113 (Figure 5). In particular, the scheduler 220 in the local storage 106 updates the statistical information 113. If t he log pointer (Log Ptr) (Figure 5) in the statistical information 113 i s set, then the local storage 106 adds the log data of this read command to the access log 114 (Figure 6). The scheduler 220 or the device driv er 219 creates the statistical information 113 and access log 114. Typi cally, the scheduler 220 is preferred in creating the statistical inform ation 113 and access log 114.

The data (requested by the read command) is then moved 707 from the cache data 212 to the host (e.g., host 104 or 105) that sent the read co mmand. In particular, the cache control 218 moves the data from the cac he data 212 to the appropriate server program (214, 215, or 216), and th e appropriate server program (214, 215, or 216) returns the data to the host that sent the read command.

Write process

In the write procedure, the cached device (local storage 106) acts a s a cache. The host 104 and/or host 105 access the local storage 106, s ince the local storage 106 has a large space. At first, a host (e.g., h ost 104 or host 105) issues a write command of NFS, CIFS, or iSCSI proto col via LAN 108, or a write command of SCSI protocol via SAN 107. The l ocal storage 106 receives the write command via LAN I/F 204 or SAN I/F 2 03. The device driver 219 (Figure 2) handles this read command from a h ost and places this read command into the scheduler 220. The scheduler 220 then analyzes the write command and places the write command into a proper server (NFS server 214, CIFS 215, or SCSI server 216) within the server programs 213. Each server program 213 checks whether the data of the write command is in or not in the cache data 212 (Figure 2). If al l data (of the write command) is in the cache data 212, then the local s

torage 106 receives the data from the host via internal bus 202 and a ne twork interface (SAN I/F 203 or LAN I/F 204). The local storage 106 wi ll over write the data on an allocated space in the local disk 209 and c ache 212. On the other hand, if all of or part of the data (of the writ e command) is not in the cache data 212, then the local storage 106 will allocate an area (in local disk 209 and cache data 212) for the rest of data not in the cache data 212. After all of the data area is allocate d, then the local storage 106 stores data in that allocated area. When a ll data is stored in the cache data 212, then the local storage 106 stor es data in local disk 209 and sends this data to a provider site (e.g., service provider site 102).

Figure 8 is a flowchart diagram of a write process in accordance wit h an embodiment of the present invention. After the local storage 106 r eceives a write command from a host, the local storage 106 checks 801 th e volume type of the data of the write command. In particular, the sche duler 220 (Figure 2) analyzes the write command to determine the volume type of the data and to determine which server program (server 214, 215, or 216) should process (handle) the read command. The scheduler 220 lo oks at configuration table 112 (Figure 4) for the volume type and the ap propriate server program to handle the write command. If the volume typ e is a static volume, then the local storage 106 skips steps 802, 803, a nd 804, and proceeds to step 805 which is described below. If the volum e type is a cached volume, then the local storage 106 checks if the data (of the write command) is stored or not stored in cache data 212 (Figur e 2).

The local storage 106 checks 802 the cache directory 210 (Figure 2). In particular, the appropriate server (214, 215, or 216) in the server program 213 checks whether the data of the write command is allocated i n the local disk 209. In step 803, if all the data (requested by the wr

ite command) is allocated in the local disk 209, then the local storage 106 skips step 804 and proceeds to step 805. In step 803, if the data is not in the local disk 209, then local storage 106 executes step 804.

If the data (of the write command) is not allocated in the local disk 209, then the local storage 106 allocates 804 data area in both the local disk 209 and cache data 212 for the write data. In particular, the cache control 218 allocates the data area on both the local disk 209 and cache data 212.

The local storage 106 then writes 805 the write data to the local disk 209 via the cache data 212. In particular, the appropriate server program (NFS server 214, CIFS server 215, or SCSI server 216) writes the date to the local disk 209.

The local storage 106 checks 806 the volume type in which the data is written by use of the configuration table 112 (Figure 4). In particular, the server program (NFS server 214, CIFS server 215, or SCSI server 216) uses the configuration table 112 to check for the volume type. If the volume type is a synchronous area, then the local storage 106 immediately synchronizes (writes) 808 the data to a service provider site (e.g., site 101) and proceeds to step 809 which is discussed below. The synchronizer 217 (Figure 2) performs the synchronizing of the data (in the user site) to the service provider site. The cache control 218 in the local storage 106 knows where the data should be stored in the service provider site by use of the configuration table 112.

In step 806, if the volume type is an asynchronous area, then the local storage 106 updates 807 the pending data list 211 (Figure 2). In particular, the server program (server 214, 215, or 216) updates the pending data list.

The local storage 106 then updates 809 the statistical information 113 (Figure 5). In particular, the scheduler 220 updates the statistical

information 113. If the log pointer (Log Ptr) (Figure 5) in statistical information 113 is set, then the local storage 106 adds the log data of this write command to the access log 114. In particular, the scheduler 220 adds the log data of the write command to the access log 114. The write process then ends.

Synchronization Process

In the case of an asynchronous volume write (a data write to an asynchronous area), the date should be sent to a service provider site (e.g., site 102) by a schedule that is defined by the user or service provider. This schedule is, for example, defined by the value in the "Policy" entry in the configuration table 112 in Figure 4. The scheduler 220 (Figure 2) executes the synchronizer 217 (Figure 2) periodically to enable data synchronization with a service provider site. This period may be set to, for example, approximately 1.0 mili-second or 10.0 mili-seconds, depending on the system 100 setting.

Figure 9 is a flowchart diagram of synchronization process in accordance with an embodiment of the present invention. The synchronization process may be performed by the synchronizer 217 (Figure 2).

The synchronizer 217 selects 901 the head data in the pending data list 211 (Figure 2). At the first, the synchronizer 217 checks 902 the pending data list 211. If there is no data in the pending data list 211, then the synchronizer 217 ends the synchronization process. If there is one or more data (in the pending data list 211) which should be synchronized, then the synchronizer 217 executes steps 902-905 as described below.

The synchronizer 217 checks 903 the information of pending data list 211. If this pending data in the pending data list 211 should be synchronized, then the synchronizer 217 sends the pending data to a service p

rovider site (e.g., site 102), so that the pending data is synchronized 904 with the service provider site. The cache control 218 in the local storage 106 knows where data should be stored in the service provider site (e.g., site 102) based upon the values in the Remote ID and Remote Site in the configuration table 112 (Figure 4).

The synchronizer 217 then selects 905 the next data in the pending data list 211. The synchronizer 217 repeats steps 902 through 905 until there is no more data to select in the pending data list 211. If there is no more data to select in the pending data list 211, then the method of Figure 9 ends.

It is also within the scope of the present invention to implement a program or code that can be stored in an electronically-readable medium to permit a computer to perform any of the methods described above.

Thus, while the present invention has been described herein with reference to particular embodiments thereof, a latitude of modification, various changes and substitutions are intended in the foregoing disclosure, and it will be appreciated that in some instances some features of the invention will be employed without a corresponding use of other features without departing from the scope of the invention as set forth.

4 Brief Description of Drawings

Figure 1 is block diagram of a system in accordance with an embodiment of the present invention;

Figure 2 is a block diagram showing additional details of one embodiment of the local storage in Figure 1;

Figure 3 is a block diagram showing one example of a data layout in accordance with an embodiment of the present invention;

Figure 4 is a block diagram of an example of a configuration table in accordance with an embodiment of the present invention;

Figure 5 is a block diagram of an example of statistical information in accordance with an embodiment of the present invention;

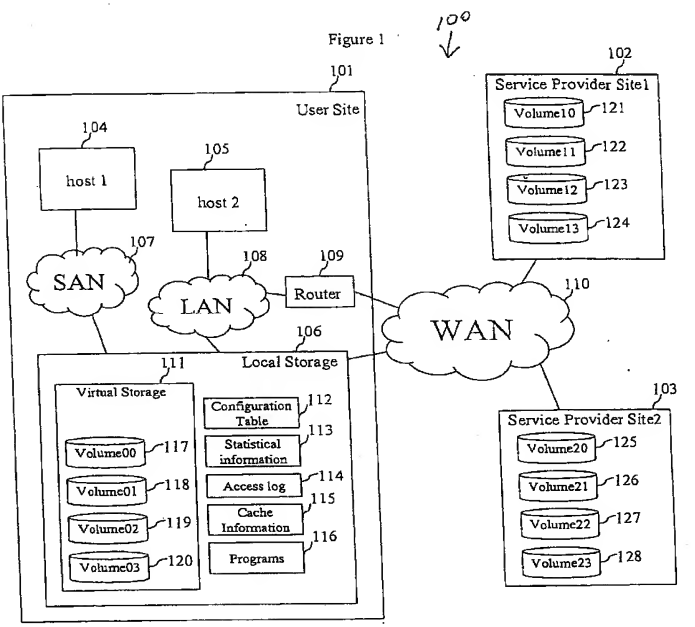
Figure 6 is a block diagram of an example of an access log in accordance with an embodiment of the present invention;

Figure 7 is a flowchart diagram of a read process in accordance with an embodiment of the present invention;

Figure 8 is a flowchart diagram of a write process in accordance with an embodiment of the present invention; and

Figure 9 is a flowchart diagram of synchronization process in accordance with an embodiment of the present invention.

Figure 1



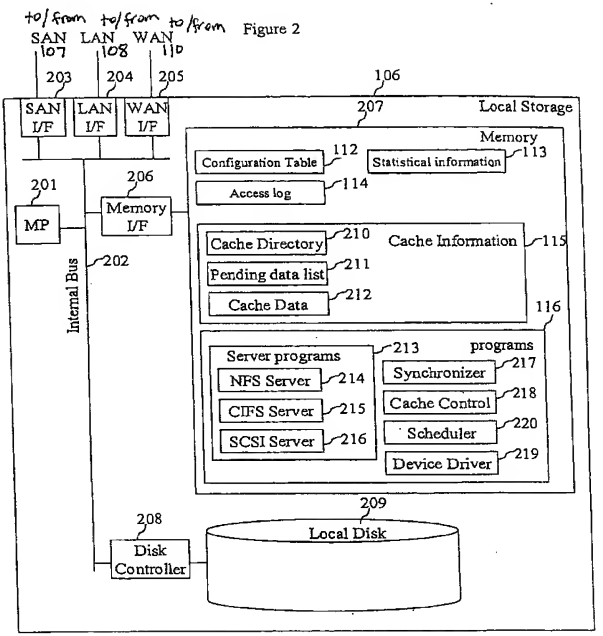

to/from
SAN
107
to/from
LAN
108
to/from
WAN
110
Figure 2
203
204
205
106
SAN I/F
LAN I/F
WAN I/F
207
Local Storage
Memory
Configuration Table
112
Statistical information
113
Access log
114
201
206
MP
Memory I/F
Cache Directory
210
Cache Information
115
Pending data list
211
Internal Bus
202
Cache Data
212
116
Server programs
213
programs
NFS Server
214
Synchronizer
217
CIFS Server
215
Cache Control
218
SCSI Server
216
Scheduler
220
Device Driver
219
209
208
Local Disk
Disk Controller

Figure 3

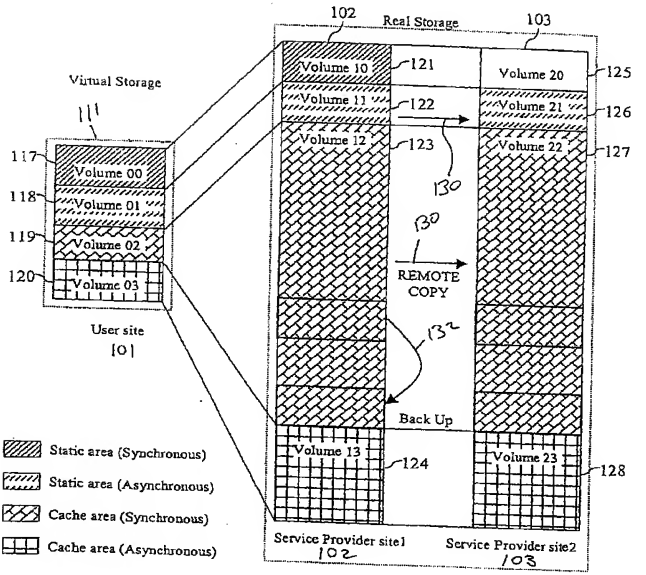


Figure 4

Configuration Table

112

ID	Interface	Sub Area	Remote Site	Remote ID	Volume Type	Service		size(GB)			Policy
						Back up	Analyze	Total	Used	Free	
00	SCSI	None	1	10	Static	Remote Copy	Null	20			null
01	SCSI	None	1	11	Static (Async)	None	Ptr	20			null
02	NFS	3	1	12	Cache (Async)	Remote Copy	Ptr	100	50	50	week
03	CIFS	None	1	13	Cache	None	Ptr	50	30	20	day

410
400
405

Figure 5

Statistical information

113

ID	Sub Area	IO/s				MB/s				Hit Ratio	Log Ptr
		Read		Write		Read		Write			
		Ave	Max	Ave	Max	Ave	Max	Ave	Max		
00											
01		15	100	3	25	0.12	2.0	0.02	0.5	80%	Null
02	/usra										ptr
	/usrb										Null
	/usrc										Null

500

510 505

Figure 6

Access log 114

Date	Time	Command	File ID	Address	Size
2000.8.1	0:0:0	Read	0	0x0000	0x0001
2000.8.1	0:0:0	Write	0	0x0000	0x0001
2000.8.1	0:0:1	Read	10	0x0100	0x0100
2000.8.1	0:0:1	Read	10	0x0200	0x0100
2000.8.1	0:0:2	Read	10	0x0200	0x0100
2000.8.1	0:0:2	Read	10	0x0300	0x0100
2000.8.1	0:0:2	Read	10	0x0400	0x0100

Figure 7

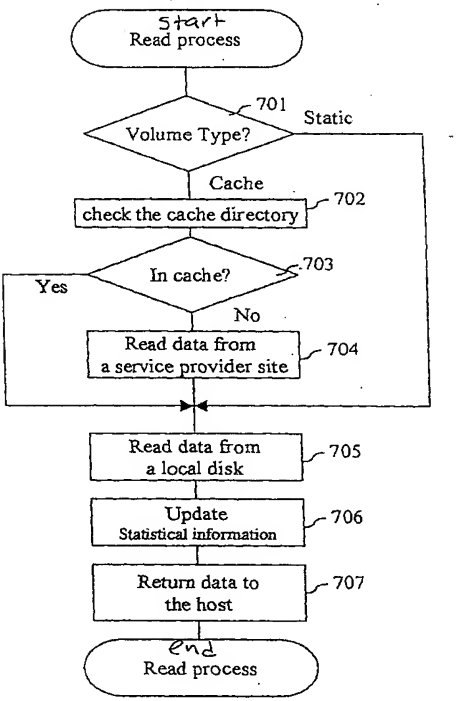

Start
Read process
701
Volume Type?
Static
Cache
check the cache directory
702
In cache?
703
Yes
No
Read data from
a service provider site
704
Read data from
a local disk
705
Update
Statistical information
706
Return data to
the host
707
End
Read process

Figure 8

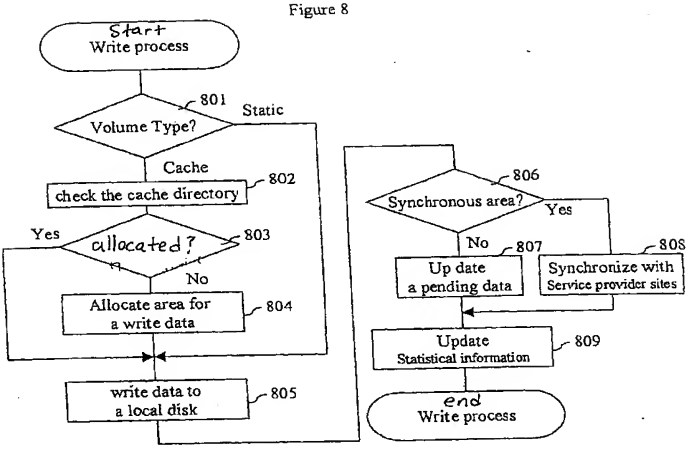
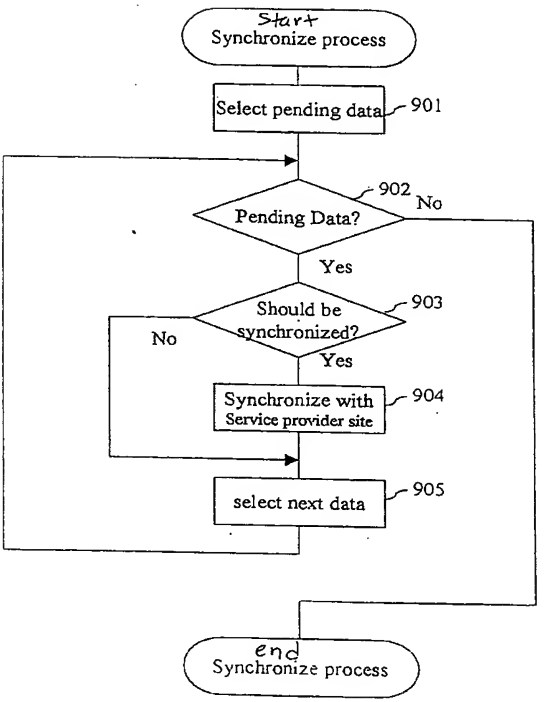

Start
Write process
801
Volume Type?
Static
Cache
802
check the cache directory
Yes
allocated?
803
No
Allocate area for
a write data
804
write data to
a local disk
805
806
Synchronous area?
Yes
No
807
Up date
a pending data
808
Synchronize with
Service provider sites
Update
Statistical information
809
end
Write process

Figure 9


Start
Synchronize process
Select pending data
901
902
Pending Data?
No
Yes
Should be synchronized?
903
No
Yes
Synchronize with Service provider site
904
select next data
905
end
Synchronize process

1. Abstract

A system for providing a data storage service, comprises: a service provider site configured to provide a data storage service; and a user site coupled by a wide area network (WAN) to the service provider site, the user site comprising a local storage having a virtual storage, the virtual storage having a synchronous volume and an asynchronous volume, the local storage configured to immediately transmit to the service provider site data that is written in the synchronous volume, to transmit at a predetermined schedule to the service provider site data that is written in the asynchronous volume, and to read data from the service provider site if the data is not stored in the local storage.

2. Representative Drawing

Fig. 1